VOLUME XLVI MAY 1955 PART 2

LEARNING AND RETAINING VERBAL MATERIAL

BY HARRY KAY

Institute of Experimental Psychology, University of Oxford

I. *Introduction* (p. 81). II. *Aims of the experiment* (pp. 81–82). III. *Experimental procedure* (p. 82). IV. *Subjects* (p. 83). V. *Treatment of data* (pp. 83–85). VI. *Results* (pp. 85–92). VII. *Discussion* (pp. 92–98): (i) *A Comparison of measures. Content: verbal and qualitative analysis* (pp. 92–93); (ii) *The relation of one reproduction to another* (pp. 93–94); (iii) *Relation of first reproduction to the original* (pp. 94–95); (iv) *The accuracy of initial recall—a comparison with Bartlett's findings* (pp. 95–97); (v) *The importance of initial responses in learning* (pp. 97–98). VIII. *Summary* (pp. 98–99). *Appendices* A *and* B (pp. 99–100). *References* (p. 100).

I. INTRODUCTION

In a series of experiments studying the formation and holding together of behavioural sequences (Kay, 1951, 1954), one of the major difficulties experienced by adult subjects was that of amending the mistakes which they themselves had introduced into their learning, so much so that in many instances the feature of the whole learning process was this unlearning of wrongly acquired responses. The following experiment was one of the attempts to examine the generality of these results by seeing if they applied in the verbal field, the most widely used of human communication systems.

Both Bartlett's (1932) technique and findings have considerably influenced the present study and, in particular, his observations upon the inaccuracy of reproduction and the persistency of a 'general form' once a first version had been given (*op. cit.* p. 93). The writer's previous experiments had confirmed this rapid establishment of a stable form which, though often far from accurate, was persistently maintained. In studying remembering Bartlett was not concerned with the difficulty of amending the errors in the 'schematized' version; in studying learning this was, for the writer, a major interest, and this experiment was designed to throw some light on the problem.

The question which was put then was in the form, 'Why should the initial reproduction be so stable when that version itself is an inaccurate reproduction of the original?' or alternatively, 'Why can accuracy of reproduction be achieved between one inaccurate version and another and not between the original and its first reproduction?'

There are many answers to this, and many of them have been given by Bartlett, but the question can be reopened by examining in some detail how far the first version is and is not an accurate reproduction of the original. Such an inquiry will not consider again the role of individual interests, attitudes, etc., in determining which parts of the original shall be accepted or ignored. Rather it will attempt to be more exact, within the limits of its own material, about the relationship between the original and the first reproduction.

II. AIMS OF THE EXPERIMENT

One aim then of the present experiment was to make a detailed investigation of the first reproduction, and consider its relation to the original. The design for this part of the experiment was straightforward enough: prose passages were read over to subjects, and after a few minutes they were asked to reproduce them in writing.

This brought the experiment to the stage where it could then study the second issue already mentioned—in what ways were subjects in their subsequent learning amending the errors in their first reproductions. The slight change from customary procedures of learning experiments which this involved will now be described.

III. Experimental procedure

Two prose passages were each read out twice to a group of subjects. Five minutes later, after listening to further material, subjects were asked to reproduce from the first *two* passages the one which they thought they could best remember. When they had finished writing this, they were asked to reproduce the non-preferred passage. After completing both, the reproductions were collected, and then the two passages were again read, this time once only. This marked the end of the procedure for that week.

One week later they were again asked to reproduce both passages and after doing this, the experimental session once more ended with the originals being read once to them. Another week later this routine was again followed and the experiment was conducted on these lines until six reproductions in all had been made. One slight departure had to be made—a fortnight, not a week, unavoidably intervened between the third and fourth versions. After the sixth version the students were told that was the end of the repeated reproduction series; the vacation followed and, at the end of the next term, four months from their last (sixth) effort, they were asked if they would attempt another reproduction, the seventh. This terminated the experiment.

The procedure has affinities with both memory and learning experiments. As with memory studies, there is some time elapsing before a subject is asked for his reproduction; as with learning studies, he is again presented with the original after his own version. The aims in compromising between the two types of experiment by extending the time interval between the presentation of the original and the actual reproduction of it were:

(1) On the one hand, to make conditions far from optimal that one reproduction should agree with another. (Both the long time interval and the interposing of the original were against this.)

(2) On the other, to weaken the effects of the original upon subsequent versions. Obviously the reproduction should follow much closer in time to the presentation of the original if the two were to agree.

Thus the wide time gap should, as compared with the normal interval in learning experiments, increase the inaccuracy both between one version and the next, and between one version and the original. These conditions for slow learning should facilitate the study of the relative influences and strength of the two versions—the original and the reproduction—upon subsequent reproductions.

The method, which is a slight variation on normal procedures, may be novel to experimental work, but it is certainly not novel to everyday life, for it is nearly identical with the school practice of giving instruction in a subject, then making some examination of what has been learned, correcting it, followed by further testing at some subsequent date.

IV. SUBJECTS

The subjects were attending a class in psychology at the Cambridgeshire Technical College (in 1950). The results of ten subjects, six men and four women aged between 18 and 35 years, are treated here. These subjects were able to complete six repeated reproductions and a final version four months later. As would be expected, an experiment (and a course in psychology) stretching over six months lost some of its starters, whilst many were not able to attend every session and therefore had gaps in their successive reproductions. Altogether results were collected from twenty-five subjects—the other fifteen have not been included in the analysed results because they have not seven successive reproductions. This was the only basis for their rejection, but a careful inspection of their records revealed the same trends as were found with the present experimental group.

When the class began, there were in it several American service personnel. The experiment was therefore planned with British and American subjects in mind, one of the two passages being in 'American-style journalism'. Owing to changes in personnel, none of the American subjects was able to complete all seven reproductions: hence the analysed results are all from British subjects and the original intention of making a detailed comparison between subjects reproducing a familiar and unfamiliar prose style had to be abandoned. Some of the broader differences between the versions of the two nationalities are briefly indicated in the results.

Material

The two selected passages were taken from Mr G. C. Grindley's Practical Class at Cambridge. One with the title 'Tea Leaves' (passage 'A') was an everyday piece of English prose, slightly and slyly humorous, describing the difficulties of selling tea. It contained 167 words, and was read first.

The second passage ('B'), entitled 'The Stenographer' was supposed to be a brash American advertisement, with short and sudden dramatic statements. It aimed to be funny and 'smart'. It contained 134 words.

Both passages are reproduced as they were read to subjects in Appendix A.

V. TREATMENT OF DATA

Verbal analysis

There are many difficulties in trying to compare one prose passage against another. As with a work of art, certain broad divisions may be generally approved but the more subtle and individual characteristics—particularly those of style—find no unanimous acceptance. It is therefore not unexpected that quantitative and qualitative evaluations of prose reproductions have often been irreconcilable. The experimenter is faced at one extreme with the problem of collecting a mass of quantitative data of which he feels, or should feel, considerable doubt, and at the other with the recording of his subjective impressions with little chance to indicate their generality and frequency.

In this experiment, for example, one subject in trying to reproduce the phrase 'the door of a speakeasy' gave it for four consecutive weeks, in spite of hearing the correct original after each version, as 'the door of a speak-quickly'. The effect on a reader, however, may be disproportionate to its frequency. On the other hand, the obvious quanti-

tative units—sentences, phrases, words or even syllables—are not interrelated on a scale of measurement, and the choice of any one may in part determine the result.

It would seem that there is no one solution to this problem but that the answer is dependent upon the length, style, subject-matter, etc., of a passage. The treatment of this experimental material was empirically determined from pilot studies with a wide range of passages. Scoring was based on word counts; there were four categories:

Correct—words reproduced accurately.

Omissions—words in the original not reproduced.

Errors—the committed errors: words substituted in the reproduction for those in the original.

Additions—words in a reproduction for which there was no corresponding word in the original.

(A further category of 'misplaced words' was considered and found unnecessary with this length of passage.)

It was felt that the four categories did provide a fair overall impression of the accuracy of a particular version. The word count is not ideal since it disregards the interrelated character of prose and treats all units as if they were of equal difficulty and significance. But these or similar objections may be laid against other possible measures, particularly the popular method based on phrase divisions, whilst the word count has the advantage of being the most objective procedure.

Content analysis

For this second measure, which was independent of the word counts, the experimenter split the content of the passages into ten sections. Three independent judges did the same, and agreement was reached on the ten major divisions. (These are given in Appendix B.) Each reproduction was then examined to see how far and in what ways its content did comply with that of the original. For each of the ten sections a subject was scored either right, wrong, or omitted. The exact fidelity of the vocabulary was not important here.

Repetition analysis

The verbal and content scores were used both in considering the accuracy of the first reproduction and in clarifying the second major issue—the difficulty of amending errors in subsequent reproductions. In a further effort to throw light on this difficulty several counts were made on the number of repetitions in a reproduction. These included for any one subject how far any one version was repeating the words and content of the *immediately* preceding reproduction, and also how far such words and content had been repeated in *all* preceding productions. Details of how these analyses were carried out will be given in the section on results.

Qualitative findings

It was intended that the objective evidence from the content and verbal analysis should be considered in conjunction with more qualitative judgements, which it was hoped they might substantiate. The qualitative treatment of the data was on traditional lines but one further examination was conducted. The seven reproductions of each subject were typed out and given to an independent judge. Then the ten reproductions for each week were randomized, and the judge was asked to assign them to their respective authors, on

the basis that there was some similarity between the successive reproductions of the same subject. This was done for all six sets of reproductions—the second to the seventh inclusive—making a total of sixty choices in all for each judge. The selections and the reasons for them were recorded from the judges.

VI. RESULTS

Verbal analysis

The data based on the scores for all subjects for both passages are set out in Tables 1A and B. The figures show the mean number of words in each reproduction and the analysis of this total into 'correct', 'errors', 'additions' and 'omissions'. The standard deviations are small enough for the mean scores to be taken as indicating the general trend.

Table 1 A. *Verbal analysis of passage A*

(Mean number of words reproduced in each version and the mean number repeated from one reproduction to the next.)

Versions ... 1		2		3		4		5		6		7	
Correct	M	M	R	M	R	M	R	M	R	M	R	M	R
Mean	44	44	32·3	53	35·6	56	45·8	62	49·8	67	56·0	53·8	47·0
S.D.	11·3	13·8	12·8	18·5	15·1	16·7	19·9	16·4	18·5	17·4	17·6	21·5	18·8
Errors													
Mean	48	48	25·2	44	27·9	45	32·4	43	31·9	40	31·0	46·2	30·1
S.D.	8·4	5·8	5·5	7·5	5·0	3·6	4·2	6·0	5·0	8·2	7·0	5·5	7·4
Additions													
Mean	21	17	6·1	22	6·5	17	8·4	16	7·0	16	7·0	21·2	8·1
S.D.	14·6	11·9	3·4	12·1	4·4	8·4	5·8	5·2	3·7	7·1	3·8	9·4	6·6
Omissions													
Mean	74	75	52·6	69	53·7	66	54·1	62	53·0	60	52·7	65·7	52·8
S.D.	11·6	13·6	12·0	14·3	7·7	15·6	7·5	15·8	12·9	17·6	14·5	19·8	16·2
Total words													
Mean	113	109	63·8	120	70·0	118	86·6	120	88·7	123	93·9	121·0	85·2
S.D.	22·5	22·4	13·3	14·4	13·6	21·3	17·9	14·4	18·5	18·9	16·5	14·8	16·7

Columns M give the mean number of words reproduced in each version, columns R give the mean number of words repeated from the immediately preceding version.

Table 1 B. *Verbal analysis of passage B*

(Mean number of words reproduced in each version and the mean number repeated from one reproduction to the next.)

Versions ... 1		2		3		4		5		6		7	
Correct	M	M	R	M	R	M	R	M	R	M	R	M	R
Mean	54	52	42·1	59	46·2	55	50·7	66	52·8	71	63·2	57	53·9
S.D.	9·6	14·6	11·4	18·8	17·8	14·8	17·6	11·5	15·2	12·1	12·2	14·6	13·3
Errors													
Mean	37	38	27·2	35	26·8	37	29·8	35	29·3	32	28·8	35	27·8
S.D.	10·0	8·6	5·9	10·0	8·5	9·3	8·2	9·1	8·4	8·5	8·5	7·6	6·3
Additions													
Mean	20	20	8·3	22	8·7	17	8·1	21	9·2	17	11·1	15	8·1
S.D.	8·1	8·1	3·9	7·6	2·9	6·8	3·2	8·4	3·4	5·7	3·9	5·3	4·0
Omissions													
Mean	42	42	29·2	40	33·2	42	30·7	35	31·8	31	30·0	41	27·8
S.D.	10·0	9·0	5·3	8·1	7·7	10·0	4·9	4·4	3·7	5·6	5·0	7·0	4·6
Total words													
Mean	112	112	77·6	116	81·7	109	88·6	120	91·3	120	103·1	108	89·8
S.D.	15·4	8·2	13·4	15·8	17·5	12·4	13·8	7·1	8·4	5·9	6·1	5·3	6·2

Columns M give the mean number of words reproduced in each version, columns R give the mean number of words repeated from the immediately preceding version.

It will be seen that the mean total number of words did not substantially change, the increase for both passages throughout all the reproductions being less than 10%. The mean number of words correctly reproduced was slowly rising until in passage 'A' it represented a 50% and in 'B' a 30% increase. But this increase was not at the expense of any one of the other three categories, which all showed some decrease. This can be seen clearly from the histograms in Fig. 1. If for each reproduction the categories are ranked according to their magnitude the only difference between the rankings for the first and sixth versions is that the 'corrects' now occupy a higher position; the others retain their rank orders with surprisingly little change in their relative magnitudes.

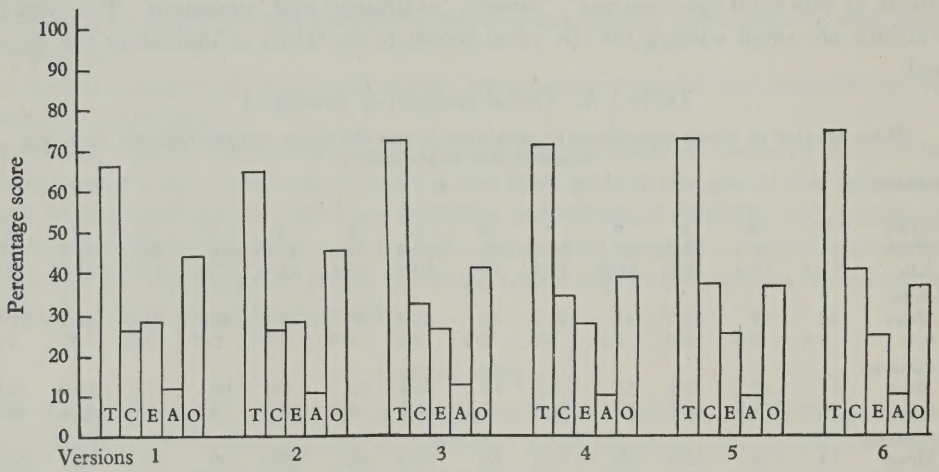

Fig. 1. Verbal analysis of the six reproductions. Passage A, percentage scores; T, total number of words; C, correct number of words; E, errors; A, additions; O, omissions.

A popular idea of learning would be that mistakes are righted, but it would seem from the above analysis that the accuracy of the passage improved partly because there were fewer omissions, partly because there were slightly fewer additions and partly because some errors were rectified. The relative improvement in each category was slight compared both with the overall improvement, and with the still remaining mistakes.

Since the passages were of different lengths—167 and 134 words—Table 2 compares the verbal scores by expressing the results for each version as a percentage of the actual number of words in that passage. The table brings out clearly that relatively more of the words of passage 'B' were being correctly reproduced and that there were also fewer omissions; to this extent the verbal findings accord with those for the content of the two passages.

Verbal repetition

The number of mistakes which occurred at each reproduction made it probable that many were being repeated from one version to the next. A measurement of this repetitive tendency has been attempted in the 'R' columns of Tables 1 A and B. This gives for each reproduction (in columns 'R') the number of words which have been repeated from the immediately preceding version. If, for example, in passage 'A' we consider the 123 total words reproduced in version 6, 94 of these were repeated from version 5. Of these 94 words, 56 out of the 67 correct words were already reproduced in version 5, as were also 31 out of

the 40 errors and 7 out of the 16 additions. Even this does not end the 'repeat story'; there were 60 omissions in version 6 and of them 53 were also omitted in the previous reproduction. It is fairly obvious that one version was bearing a much greater semblance to its immediate predecessor than either was bearing to the original.

Table 2. *Comparison of the verbal scores in the two passages*

(The means of Table 1 expressed as percentages of the actual number of words in each passage.)

Version ...	1	2	3	4	5	6	7
Correct							
Passage A	26	26	32	34	37	40	33
Passage B	40	39	44	41	49	53	42
Errors							
Passage A	28	28	26	27	25	24	27
Passage B	28	29	26	28	26	24	26
Additions							
Passage A	12	10	13	10	10	10	12
Passage B	15	15	17	13	16	13	11
Omissions							
Passage A	44	45	41	39	37	36	39
Passage B	31	31	30	31	26	23	30
Total words reproduced							
Passage A	66	64	72	71	72	74	72
Passage B	83	83	87	81	90	90	80

Table 3. *Verbal repetition from all preceding reproductions*

Versions ...	1		2		3		4		5		6		7	
Passage A	Mean	%	Mean	%	Mean	%	Mean	%	Mean	%	Mean	%	Mean	%
Correct	—	—	32	73	30	93	29	97	28·5	98	28	98	27	96
Errors	—	—	25	52	21	84	18·5	88	17	91	16	94	15	93
Additions	—	—	6	35	4	66	3	75	2·5	83	2	80	2	100
Omissions	—	—	52·5	71	47·5	90	45·5	95	44	96	43	97	40	93
Passage B														
Correct	—	—	42	77	40	95	39	97	38·5	98	38	98	37	97
Errors	—	—	27	73	23	85	22	95	21	95	20·5	97	19	92
Additions	—	—	8	40	6	75	4·5	75	3·5	77	3	85	2·5	83
Omissions	—	—	29	69	26	90	24	92	23·5	97	23	97	22	95

Each score gives the mean number of words in that reproduction repeated from *all* preceding reproductions. Percentage scores express the above mean scores as percentages of the total possible scores for that reproduction.

Table 1 has illustrated the prevalence of a certain kind of word repetition—that from the immediately preceding reproduction—but it now seemed relevant to ask how frequently was a word repeated when it was correct and when it was wrong. There was the further question of the probability after a mistake had *often* occurred of it being repeated, and whether it was more likely to be reproduced again than a mistake which had been made only once. Table 3 attempts to answer some of these points; it details the mean number of words in each version which have been reproduced in *all* the preceding versions. Thus if we are considering the errors in version 5 (passage B) the table is only concerned with the 22 errors which have already been reproduced in versions 1–4: of these 22 errors 21 were again made in version 5.

This type of measure is of course very strict, for if a word drops out only once in any reproduction it is thereafter lost in all the measurements of subsequent versions. It will be appreciated that this implies that in Table 3 in any one category the total *possible*

score for one reproduction is the actual score of the immediately preceding version; and similarly that score, by the same definition, could not be greater than that of any of its preceding versions. Hence almost inevitably there is some decrease in the means from one version to the next. But of the words which do comply with these requirements a very high proportion are subsequently reproduced. This is illustrated in those columns of Table 3 which express the results as a percentage of the total possible score for that reproduction. In the above example, 21 errors were repeated of a possible 22, that is 95%.

The results expressed in the table seem definite enough—if a word was reproduced three or four times then the probability that it would be reproduced in a subsequent version was extremely high. In this experiment where learning conditions were slow that probability was nearly as high for an error as for a correct word. The point also applied to words appearing in the original and which had been repeatedly omitted—there was a high percentage of words which subjects had never attempted to reproduce and the longer such words had been ignored the more likely in fact was this to happen.

It should be pointed out that the 'Additions' which were repeated, represented only a very small proportion of the total additions for most of the reproductions. The words which a subject added to a passage, seemed to vary more from one reproduction to another than do words in the other three categories—corrects, errors, and omissions. Hence, though the percentage figures for the additions were high, they were percentages of very small totals and are not reliable.

Content analysis

As already indicated, all the reproductions of every subject were examined to see how far the content agreed with that of the original.

Table 4 gives the results. If we begin by considering only the content which was actually reproduced at each version it is obvious that a very high percentage of both passages could be recalled at the first attempt. Within whatever individual framework

Table 4. *Total scores for each reproduction of content of passages A and B*

(Total possible score for each reproduction—10 items × 10 subjects = 100. Correct, C; omissions, O; errors, E.)

Versions ...	1			2			3			4			5			6			7		
	C	O	E	C	O	E	C	O	E	C	O	E	C	O	E	C	O	E	C	O	E
Passage A																					
Content actually reproduced	59	28	13	51	38	11	61	28	11	64	25	11	63	25	12	68	23	9	62	25	13
Content repeated from immediately preceding version	—	—	—	51	28	10	51	26	9	60	25	10	63	25	11	63	23	9	61	23	8
Passage B																					
Content actually reproduced	83	13	4	86	12	2	86	12	2	83	12	5	86	9	5	90	6	4	83	11	6
Content repeated from immediately preceding version	—	—	—	81	11	2	83	11	2	82	11	2	82	9	4	85	6	4	82	10	4

subjects were holding and expressing their ideas there is no doubt that most of the substance of each passage had been assimilated. The striking feature *is* that so much had been grasped from the beginning.

It would be suspected from these results that the content actually reproduced in one version would be very similar to that of another. This further analysis was carried out,

and Table 4 also reproduces the scores showing how far the content of one reproduction repeated that of the immediately preceding version. The figures do establish very definitely how little change there was in content from one version to the next. As some learning did take place there was a slight trend for more substance to be correct and less to be omitted, but the newly acquired substance was added to that which was already held whilst that which was finally omitted had never appeared.

It had been intended to establish this point by a further analysis of the content, examining how far a reproduction was repeating the substance of *all* preceding versions. But it was felt that the data of Table 4 sufficiently indicated that so far as content was concerned, subjects were repeating the same correct items, making the same errors, and omitting the same items one week after another. In other words, the initial reproduction had established the content of the passage; as it was laid down then so, with only slight modifications, was it held six weeks and ultimately six months later.

Qualitative findings

We can now turn to the more qualitative analysis of the passages. In order to obtain some indication of the individuality of a reproduction all the versions of all the subjects were typed out on separate sheets of paper; following the procedure already described (see p. 84) five independent judges were asked to select at each successive stage the six reproductions of each subject. No mistakes were made, and it was obvious that though the process was tedious there were ample cues upon which to base the assessment. (The statistical significance of the result even with five judges is of course very high.)

When the judges were asked to say upon what they had based their judgements, in addition to such points as the amount of the passage reproduced, the particular parts omitted, and the general accuracy of the ideas and the words, they also stressed the recurrence of key words and phrases, and the ease with which certain individual stylistic features could be recognized. It is hoped to illustrate some of these features in the following examples.

Minor but persistent verbal changes

In this type of example the verbal alterations were of a very minor order incurring little or no change in meaning. This often makes it the more interesting because such 'error' words could have been so easily interchanged. It was also noticeable that they occurred in sections where much was accurately reproduced. A typical example was where a subject, no. 1, in all his seven versions wrote for 'the actual political opinions of mankind' the phrase 'the political ideas of man'. As a change in meaning it is negligible but as a change that can persist over six months it is of note. Similarly, the same subject, again in all his seven reproductions, uses 'the entire stock' for 'the whole stock', a 'vague urge' for 'a vague impulse' (the word 'urge' was frequently used by other subjects in this context) and 'a picture of the Tropics' for a 'vague suggestion of the Tropical East'. Another subject, no. 3, who had been in the East, in his seven reproductions always referred to the tea as growing 'on a bush' and in three of these it was 'on a bush in Ceylon'. Subject no. 7, a grocer by trade, instead of the phrase 'What therefore should the advertiser do to create a commercial entity', always gave 'What is the *purveyor* of tea to do to bring his product'. For the same subject the 'subconscious memory of a geography lesson' was always 'a half forgotten geography lesson'.

The same persistence in adhering to minor adjustments of the text was found in the reproduction of passage B. The phrase 'instead of throwing tear-gas bombs' caused many errors; one subject, no. 8, in all his seven versions gave 'faster than tear gas' whilst subject no. 9, apparently economically minded, offered in his first five versions 'saving the police using tear gas' and then in his last two versions extended this to 'saving the police the expense of using tear gas'.

In the last sentence of the passage the phrase 'to have the office heavily insured' was usually reproduced more elaborately, as in the six repeated versions 'should ascertain he has a full fire insurance'. Several subjects apparently seemed to think they were catching the American idiom by following the popular principle of never saying in one syllable what could be said in two. It may have been that certain phrases seemed to possess some American quality which influenced subjects in adhering so repeatedly to them: no. 2 maintained his incorrect 'city police' for five reproductions only to change it in favour of the more inaccurate 'chief of the city police'; subject no. 5 was equally insistent on repeating 'steel framed equipment'.

As was observed at the beginning of this section, it is not the magnitude but the persistence of such deviation which is the relevant point. It does seem surprising that at weekly intervals, and sometimes longer, anyone should repeat so exactly what they have previously reproduced when their last perception was not of their reproduction but of the original which corrected their errors.

Some major verbal changes

The examples in this section were often more noticeable than the types so far discussed, and perhaps for this reason they were somewhat less persistent.

Some of the changes, without introducing any different meaning, were striking in their new phraseology as when one subject (no. 8) in six consecutive reproductions referred to 'a name with an Eastern flavour'. Subject no. 10 would persist for five reproductions in using the phrase 'man in the street' (which appears in neither original) and even included it twice in two reproductions. The phrase dropped out of the sixth reproduction but returned after the four months' interval in the seventh. An interesting embellishment in which several subjects indulged was to make a 'geography lesson' into a 'school geography lesson'. Here perhaps we are not dealing so much with a change in phraseology as a persistent way in thinking—namely that geography lessons are usually (or ought to be?) school lessons.

The type of alteration which did involve a change in meaning was that by subject no. 5 who amalgamated 'the geography lesson on Australia' with the phrase 'the recollection of his grandmother' to make for five versions 'the grandmother in Australia', whilst subject no. 3 perhaps influenced by popular notions of grandmothers and tea drinking—there used to be many pictures on tea caddies of old ladies holding teapots—three times reproduced 'the recollection of Granny's teapot'. The strength of this tendency to incorporate general ideas into specific material may be gauged by the habit persisting even under these conditions of reproducing and then presenting the original material at weekly time intervals; presumably this type of mistake would be more prevalent if the intervals were longer.

In passage B some of the omissions brought out how easily phrases were dropped out

when they did not advance the narrative. The unnecessary 'into the building' was continuously omitted by nearly all subjects, and the description 'in front of the door' was rendered by the majority as 'outside the door'. It was perhaps because of a lack of familiarity with the idea of a police 'raiding squad' that everyone referred to 'the police' but many omitted 'raiding squad'. It was also noticeable that most subjects began by saying 'I type quicker than a new band', etc., and omitted any reference to 'take shorthand notes'; it was as if typing had been the main subject-matter of the passage and had pushed out any reference to the other quality in spite of it occurring at the beginning of the original. There was the usual persistence on a number preference, one subject insisting on interpreting 'several large firms' as 'the last three firms' and in two versions she also 'set fire within three days' to the offices, whilst another subject would repeat 'three weeks' for 'a few days'.

The effort to reproduce 'Style'

So far examples have quoted short phrases and brought out alterations in individual words. And even when illustrating qualitative results it is difficult to do more than this. But reading through the reproductions it was obvious that a subject had reacted to much more than we have so far conveyed. The words were merely the hard structure of each creation; within this verbal framework not only had a set of ideas—the main content of the passage—been more or less accurately conveyed but they had been transmitted with very obvious awareness of the stylistic features of their presentation. The reactions to the different styles of the two passages was most emphasized. At the same time the attempts to reproduce that style were often anything but successful, and as already indicated, the styles of any one subject's reproductions tended to bear more relation to each other than any one of them did to the original.

This point was most clearly brought out to the experimenter when comparing the versions of passage 'B' by the American and British subjects. Though in many ways the mistakes of both nationalities were the same, and along the lines already indicated, there were stylistic differences, notably that the American subjects had extended the ideas, and sometimes the jargon, of the original. (The American subjects without exception had chosen to reproduce 'B' before, and at greater length, than passage 'A'—the British had done the reverse.) As examples of their reproductions, the bootleggers became 'a couple of Prohibitionist bootleggers' for one subject in all his four versions; another maintained it was the 'Oklahoma City Police Vice Squad'; a third made it 'F.B.I. officials'; a fourth described it as 'the Police riot squad'; whilst a fifth related how 'the State Police hired me to sit on the side walk'. The American idiom is obvious enough here as it was with the only subject who put the passage into the third person and began 'A small town in Oklahoma...' and finished all four versions with 'She warned her next employer to watch his step as he was going to get quite a girl'. (The three American subjects who gave any indication on this point all assumed the stenographer to be feminine.)

As a very brief summary of the qualitative differences between the versions of the American and British subjects in this passage, it may be said that the American reproductions were elaborated, almost casually and inevitably, by a familiarity with a wealth of background material. When a subject moved outside the limits of the passage he added local colour to it. It would be difficult to distinguish how far content and how far

particular idioms and phrasing were contributing to this total impression. But the result was that many of their versions had lost the pseudo style of the original and sounded authentic. British subjects, on the other hand, when they departed from their text, were often too obviously striving to achieve a style with the result that their versions were more uneven. They could be consistent with themselves, but not with the original.

The total impression given by the reproductions of the two nationalities was very different, but their versions established clearly enough what were the main reactions to the passage—a generally accurate construction of what it has said and at the same time a definite awareness of how it had said it. These together made up the passage, but matter and manner were modified against the background of the individual's experience, and the process seemed as recognizable under these conditions of learning as it was under the customary Bartlettian Remembering Studies.

VII. Discussion

(i) *A comparison of measures. Content: verbal and qualitative analysis*

The relationship between the results of the content and verbal analysis has been graphed in Fig. 2. Whilst approximately 60–70% of the content was accurately reproduced in most versions, the percentages of correct words was only half of this, lying between 25 and 40%. Conversely, only a small percentage of the content was wrongly reproduced—9–13%—whilst the verbal errors for all reproductions was around 25%. The differences between the omissions are not so marked but are still definite enough—for the content they lie mainly between 25 and 28%, and for the verbal from 36 to 45%.

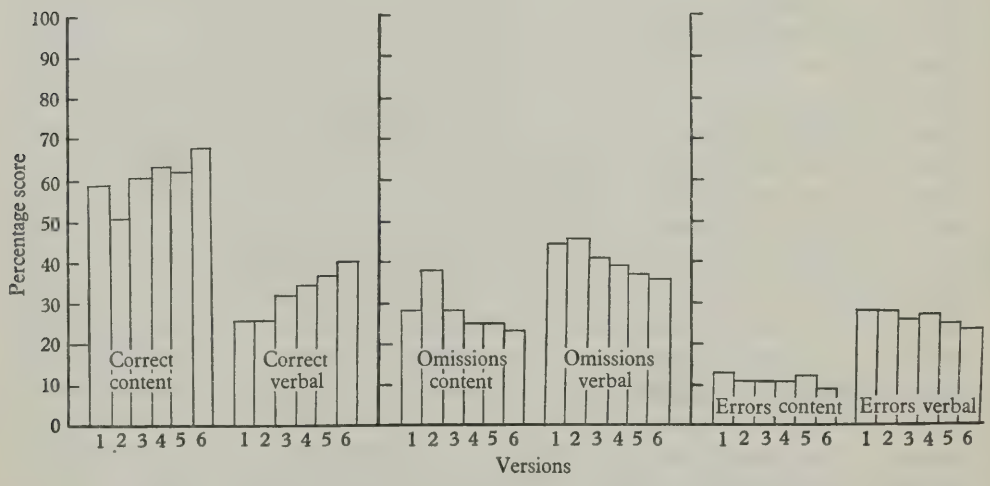

Fig. 2. Comparison of the content and verbal analysis. Percentage scores.

The discrepancy between the two scores raises the problem as to how the accuracy of a reproduction should be assessed. Here if only the content is considered, then it is a fairly satisfactory rendering of what has been said; but if the actual vocabulary of the original is required, then it is not.

It is fairly obvious that no inflexible rule can be laid down for making a judgement between these two criteria. If the passage is a short one, then the accuracy of the verbal reproduction will improve in comparison to the content analysis; with increasing length it will probably suffer. But this is supposing that content and vocabulary remain within one range of difficulty. Where, for example, the content becomes complex or abstract, then its reproduction is likely to suffer accordingly—some items in the 'A' passage which were consistently reproduced inadequately in the present series illustrated this point. On the basis of the present results it can only be argued that both measures should be considered if anything like an overall impression of a version is to be achieved.

In point of fact this has not often been done; it would follow that where results of different verbal experiments are being considered, it is essential to know the criterion upon which they are based. Generally more weight has been attached to the qualitative observations. It would seem almost inevitable that such observations would magnify differences between a reproduction and its original—where the discrepancies are striking, the reader tends to be over-influenced by them and does not allow sufficiently for the relative infrequency of such examples. For this reason the verbal analysis is providing useful supplementary evidence to the more qualitative demonstrations of verbal changes. The two may be regarded as complimentary—the analysis giving the frequency and the qualitative examples indicating the magnitude of such errors. In spite of the doubts which were expressed about the verbal analysis, it is felt that with this length of passage it can serve a useful function; at the same time it is the study of the words themselves which is likely to throw most light upon what a subject was doing, and trying to do.

(ii) *The relation of one reproduction to another*

Taking into consideration all the above measures, we can now take up the question of the relation of one reproduction to another. Though most studies of serial reproduction have stressed how closely one version follows another, in the present experiment, with the interjection of the original after each reproduction, there was added opportunity for subjects to amend their errors. In a very limited way they did this; they were learning, but the outstanding result was that whatever the measure considered, each subject established an extraordinarily close relation between one reproduction and another. This applied to the form and style of the passage, to individual words and minor items, in addition to the general content and overall impression. The result was, in effect, that where a subject had mistakenly introduced errors into his reproduction for two or three weeks, though he had heard the original with its corrections of his mistakes after each of his versions, the chances were very much in favour of him not amending those mistakes but of repeating them in exactly the same form. The point also applies to words which had been omitted—some words in the original were never reproduced, and the longer they were ignored, the more likely was this to happen.

Thus, the reproductions which a subject was making were steadily becoming more rigid and predictable, in that with each succeeding attempt a reproduction was becoming more and more like its predecessor. If the original was ever to be learnt, these errors of omission and commission had to be overcome, but it was obvious that this was not an easy step, in so far as certain words, phrases and ideas, had been 'built into' these early reproductions.

It would follow from what has been said that the first reproduction was all-important in laying down the future lines upon which a subject would approach the material. Thereafter, consistency was achieved in spite of the 'intrusion' of the original. This has an obvious bearing upon learning studies which will be illustrated in the last section; meanwhile, it leads us to a consideration of the issue—the relation of the first reproduction to the original.

(iii) *Relation of the first reproduction to the original*

In answering the question 'How accurate is the initial reproduction?' Bartlett was definite enough: 'Accurate recall is the exception and not the rule' (*op. cit.* p. 61; cf. also p. 93). Bartlett, of course, is not referring specifically to the initial or any other reproduction; but his schematic theory of memory makes it obvious that he would be regarding the initial reproduction as one stage in this continuous series of dynamic changes.

With our material, Bartlett's judgements would again apply to the initial reproductions; only about 30% of the actual words in the two passages are correctly given. But the statement, whilst accurate enough within its own limitations—'in a literal sense'—would be a misleading summary of the more general features of the reproductions. Most of the content—approximately 70%—was being accurately conveyed by the initial versions; on this criterion the first reproductions are satisfactory whilst their general form and phrasing reflected a definite effort, though not necessarily successful, to convey the different styles. Essentially these reproductions represented a total reaction to the original material—total in so far as it included not only a reaction to the obvious features such as the content, but also to the more subtle such as how a thing was being said, what kind of a person might have said it, what was its background, etc. In discussing the passages with subjects after the experiment it was frequently brought out that without having formulated their reactions they had some fairly persistent impressions which were more definite than they themselves realized; they had categorized material and events without necessarily being aware that they had done so. It was only when they were asked why they were making a particular statement that they realized their previous assumptions. This could be illustrated by precise examples—the air-crew pilot still engaged in flights over Germany who four times reproduced the 'varied political opinions of mankind' as 'varied as the political opinions expressed in *Mein Kampf*'; the subjects who had 'reproduced' a feminine stenographer, and those who had assumed a sex without stating it; but over and above these particular instances there were the more generalized impressions which seemed to have permeated and be effective in determining what a subject was attempting to reproduce.

This point was clearly illustrated in another context when using *picture* material with the method of serial reproductions. After subjects had heard a description of the picture and passed it on, they were questioned about their other impressions of it; what period of history did it represent? the relative sizes of objects? could they draw it? etc. The answers were generally negative and an impression was conveyed of extreme vagueness or even blankness. Then the original was shown to them and the instantaneous effect upon a subject was almost always one of incredulity—often a gasp of astonishment. When asked why, a typical naïve reply was 'I did not imagine it to be like that'. 'But you said you couldn't imagine it.' 'No, I couldn't, but somehow that is different.'

In other words, the subject was holding more than he could convey, and what he was

holding was formed from a reaction between the limited new material he had been given and the almost unlimited resources upon which he could draw and against which he had assimilated the new. In the present passages, as with picture material, the subject was aware of what he had been given, but was unaware how far he had made a subjective reassessment of it. With the picture material this came out graphically when two subjects could agree upon a verbal description of certain details but could not agree upon the authenticity of the original picture; that is to say, they could agree upon what they themselves knew they were holding—their disagreement was arising over their holding so much more than they were able to express.

We may now ask how far these findings on the general accuracy of the initial reproduction would agree with Bartlett's exposition.

(iv) *The accuracy of initial recall—a comparison with Bartlett's findings*

Bartlett's monograph served magnificently, not only to illustrate the fallacies in everyday remembering, but to examine where and why they occurred. There is, however, the danger that the victory has been too complete. Having destroyed the myth of a verbatim accurate reproduction and seriously damaged a trace theory of memory in favour of a more loosely conceived and modifiable scheme, the question now arises: How far are memories ever accurately conveyed, and, if they are accurate, how is this achieved? Fortunately we are attempting to answer the first question, though this may throw some light on the second.

Bartlett's emphasis is always on the point that recall is a constructive process, largely determined by an attitude which is itself arising from the initial impressions left by the original (e.g. *op. cit.* pp. 176, 205, 207). 'It is this impression, rarely defined with much exactitude, which most readily persists', and the details are built into and around it so that they may satisfy the general setting.

This is, of course, a difficult point to discuss, since obviously we can only deal with such a process at second-hand through its products; here we can never say what that impression was, but, of necessity, have to base our assumption upon the results of some subsequent reproduction. And it may well be that under everyday conditions, where individuals are perceiving material without being directed by any instructions to remember it, their constructive reproduction is as imprecise and affectively determined as Bartlett has made it out to be. But under conditions such as the present, Bartlett's observations hardly seem to give sufficient weight to the accuracy with which the substance of the passage could be initially reproduced. This would apply to the results of all our subjects: in examining their first reproductions, we seem to have the product of some process whereby the material has not only been assimilated, but, to some extent has been carefully sifted and evaluated—the version which nearly all subjects reproduced did contain the main items in the passage.

The closer Bartlett's work is examined, the more would it seem that the examples which fit his account of constructive recall, and which he himself most frequently mentions, are those where subjects are trying to recall material after a long time interval—at least after some months and sometimes years. In some cases so much of what was originally given has all too obviously been 'lost'. The subject struggles to create something of the attendant circumstances—the time, place, people, etc.—anything upon which he can

hinge a specific item and he is tremendously influenced in this searching by the probability of an event. The writer studied this process by asking student-subjects the question 'Can you recall your first University lecture?' and noted their introspection. It often became impossible for a subject to distinguish whether an event had taken place because he remembered it, or because he knew it was likely to have happened at that time. With the recall of events of this order Bartlett's exposition of a constructive effort seems apt enough, but there is a pronounced difference in degree, if not in kind, between this long-term remembering and the recall of material that has only been held for a few minutes. There is obviously less searching over a range of possibilities, certain items slip out promptly whilst others are quickly grouped around them. The question arises how far is this itself a constructive process, and how far is it the product of a 'constructive' process which has already taken place—in fact, which has taken place at the time of the initial perception?

Bartlett himself provides evidence on this point for it is of course an integral part of his thesis that perceiving and remembering have the same active characteristics, both being dominated by what he has called the 'effort after meaning'. But by reiterating the importance of a constructive process at the time of recall, Bartlett does not seem to have given sufficient weight to his own earlier observations of the active nature of perception itself. If his observations are right, as they seem to be, then at the first presentation of the material this effort to extract 'meaning' has begun. Bartlett has indicated the variables which influence the procedure, but we are concerned with the effect of this assimilation of material on this perceptual stage. To some extent passages may suffer an unwitting transformation, speech habits and verbal idiosyncrasies are incorporated into a new version according to personal predilections and social conventions. But the most significant feature is that this perceptual analysis—and in many ways this does seem to be a fair description of the process—is more accurate than is generally allowed. Whatever we may be conceiving as the retaining mechanism by which this is stored, it results in an appreciation of what has been presented, which, by overall standards, is a comprehensive digest of the material.

The argument then is stressing two points:

(1) That where material is being remembered after relatively short time intervals, it is necessary to give more weight to the interactive processes which have taken place on the initial perception of the material, and less to the constructive effort at the time of recall.

(2) In so far as the first stage of remembering—the initial perception—is an interaction directly involving the presented material, whereas the constructive stage of the final recall does not, the first is the more accurate. (It is assumed here that retention is a dynamic process, involving the possible modification of remembered data. Hence, a subject in trying to recall material is not necessarily holding the identical material which he began by retaining.) In view of the nature of this initial interaction, it seems somewhat misleading to speak of it as a 'constructive' process, even if we are prepared to apply this term to the recall stage of remembering. Accordingly, in an effort to place more emphasis upon the presented material it is proposed to describe this initial assimilation, not as a 'constructive process' but as a 'perceptual analysis'. It may well be that under particular conditions this analysis is defective—where, for example, the material is abstruse or excessively lengthy or disconnected so that items may easily be omitted. Under normal conditions it is not *literal* recall, but its deficiencies, though striking, should

not obscure the fact that it successfully assimilates a mass of material that would be well beyond the immediate memory span.

Since completing this experiment the writer has had the opportunity to read and discuss the research which Gomulicki, working under Zangwill, has done upon 'Recall as an Abstractive Process' (Gomulicki, 1952). As this is now being published the only comment which should be made here is that the operation which he has described as an 'unwitting abstractive process' seems in most respects identical with the stage which we have called a perceptual analysis of the material. His term seems preferable and would be adopted by the writer.

It may be noted that the proposal to place less emphasis upon the constructive stage of recall, and more upon the original analysis which occurred with the first presentation, offers some partial explanation for the consistency of one reproduction with another under the conditions of the present experiment. At each stage of reproduction the important variable would not be the reconstructive process peculiar to that stage alone but the initial analysis common to all recall stages. The point would also have some bearing upon the importance of the initial reaction to the material in learning studies which we may now consider.

(v) *The importance of initial responses in learning*

It was noted in the section on Experimental Procedure that the time intervals were being manipulated to encourage slow learning in order that some estimate might be made of the relative strengths of particular variables. We will now see how this did in fact operate, and briefly consider the bearing of those results upon learning studies.

In passage A there was mentioned a specific sum of money, £100,000. It would be expected that errors would occur and they did. Some subjects could repeatedly reproduce an amount consistently in an erroneous variant; one subject gave six of his seven versions as £1000 and another gave five as £50,000. One subject did have a general selection of £20,000 (twice); then £10,000 and finally, £100,000. This last subject may represent the person who for a time just did not know what was the exact sum but was aware that it involved a big amount: but the examples of subjects repeatedly giving a wrong variant would seem to illustrate a different procedure. There are several possibilities which might account for this mistake:

(1) Subjects may have been so sure that they knew that particular part of the passage that they did not check it carefully when it was read out to them each week—the mistake persisted because they were not attending to the correct version.

(2) This is harder to conceive but perhaps subjects did observe the discrepancy between their version and the original but had forgotten it by the following week.

(3) A third possibility is that some observed the original without at first being aware of the discrepancy between it and what they had previously reproduced. Subsequently, it was their own version which they retained and reproduced in the following week. This would argue that the impression of their own version was stronger and more permanent than that of the original. On this third view learning would finally take place when a subject became aware of the discrepancy between the original and his own version.

On the evidence here it is not possible to be definite. But subjects who had repeated the same mistakes and had subsequently learned the correct version were questioned, and it was found that they did know both the correct answer and their previous error. Without

claiming that the procedure of the third possibility was ubiquitous, it would be favoured as being at least prevalent. It would explain the occurrence of many of the mistakes on the same grounds as were found in the Serial Learning experiments (Kay, 1951), where complete learning was frequently taking place only after the identification of the error itself. In support of the generalization made there—learning involved unlearning—this experiment both illustrates it and extends its application. On the one hand, it applied to processes involving everyday speech habits—a word in the original was blocked by one from the subject's own vocabulary, e.g. 'urge' for 'impulse', 'purveyor' for 'seller', etc., and it was necessary for him to learn that his own word was wrong before he would replace it. On the other hand, there were broader, overall reactions to the material as a whole which a subject had made and perhaps because of their imprecision they were all the more difficult to modify; as noted in some cases subjects did not seem aware of their influence.

To conclude this section, the initial reproduction represented the interaction between what was given and what a subject brought to the task. The importance of this initial reaction was known from memory studies; it would seem that under these conditions where learning was possible its importance was no less. This was primarily because the subsequent modification which had to take place, if the material was to be learned, had to be undertaken by an organism which had already exhibited a predilection for the mistake which it had introduced. In learning it would be easy for one subject to amend the mistakes of another; it is altogether more difficult for a subject to amend his own. This experiment is in line with those which when considering the normal learning reactions of a mature organism have abandoned any S—R model and have placed the emphasis more and more upon what is already held, and less and less upon what is presented. What is initially laid down represents the interplay between 'permanent' directive tendencies within the organism and the presented material. Once that interaction has taken place the products are not easily modified. But there is a danger of overstating the case. In so far as learning generally demands complete accuracy, the difficulty of achieving it and the strength of what is already held is not misleading. But in noting the initial errors it is as well to be mindful of the general accuracy of the whole.

VIII. Summary

An experiment was carried out with verbal material based on the method of repeated reproduction. (Weekly time-intervals.) But the standard procedure was modified in that the original version was read to subjects immediately after each reproduction.

Under these conditions it was found that:

1. The initial reproduction was a fairly accurate version of the content of the passage (70%), but less so of the actual verbal material (30%). The need for considering both criteria is brought out.

2. Subsequent versions were strikingly similar to each other and in their repetitions from the initial reproduction. They reached a stereotyped form comparable to that found in normal repeated reproduction experiments. This comparability applied to both content and verbal analysis.

3. The constructive nature of recall has been stressed by Bartlett but it is felt that in short-term studies emphasis must be placed upon the initial interpretation by a subject.

This interpretation is described as a 'perceptual analysis'. The process is compared to Gomulicki's recent study of an 'abstractive' operation. His term is preferred.

4. The relevance of the experiment to learning studies is brought out. Again it illustrates the necessity for an unlearning of mistakes, and in so far as the organism has shown a predilection for these erroneous introductions—either verbal or conceptual—the difficulty of their amendment is considerable.

APPENDIX A

Passage 'A'. Tea leaves

The actual tea leaves in the world are as varied and unstable as the actual political opinions of mankind. Every tea leaf in every tea garden is different from every other leaf, and a week of damp weather may change the whole stock in any warehouse. What therefore should the advertiser do to create a commercial 'entity', a 'tea' which men can think and feel about? He would choose some term, say 'Paramatta Tea', which would produce in most men a vague suggestion of the tropical East, combined with the subconscious memory of a geography lesson on Australia. He would then proceed to create in connexion with the word an automatic picture-image having previous emotional associations of its own. By the time that a hundred thousand pounds had been cleverly spent, no one in England would be able to see the word 'Paramatta' on a parcel without a vague impulse to buy, founded on a day-dream recollection of his grandmother, or of the British Fleet.

Passage 'B'. The stenographer

Wanted stenographic position—I take shorthand notes quicker than a new band of Oklahoma outlaws can spring up. I lay a smoke screen with a typewriter, and I do not stay long with one firm because I work so fast I burn the office down.

References: several large firms in Oklahoma City had complete new furniture of steel installed after I had worked there a few days, and the police raiding squad stationed me with a typewriter, and copying the encyclopaedia, in front of the door of a speakeasy to smoke out the bootleggers instead of throwing tear gas bombs into the building. Warning —I advise the person who contemplates hiring me to have his office heavily insured, because if I work for him he will have a sure case of conflagration on hand.

APPENDIX B

Content of the two passages

Passage 'A'. Tea leaves

Item 1. Tea leaves are as varied and unstable as political opinions.
2. Every tea leaf is different from another.
3. A week's damp weather changes a warehouse's stock.
4. How can an advertiser create a saleable object evoking thought and feeling?
5. Give the tea a name—Paramatta.
6. This name would suggest the Tropical East and memories of a geography lesson on Australia.
7. Create around the name a picture-image with an emotional association.

8. Spend £100,000 cleverly.
9. All England would want to buy when they saw the name 'Paramatta'.
10. This desire found on recollection of grandmother or British Fleet.

Passage 'B'. The stenographer

Item 1. Wanted stenographic position.
2. I take shorthand notes quickly (quicker than new band of Oklahoma outlaws can spring up).
3. I type very fast (so fast I lay a smoke screen).
4. I do not stay long with firms because I set the office on fire with working so fast.
5. References: Large firms in Oklahoma City.
6. Had steel furniture installed after I had worked there.
7. The police raiding squad put me outside a speakeasy.
8. The smoke from my typewriter was used instead of tear gas bombs on the bootleggers.
9. Warning: insure your office if you hire me.
10. I shall certainly cause you a fire.

There was no point in trying here to translate the content of passage 'B' into 'English'. But in marking the content of both reproductions no attention was paid to the literal accuracy of words.

References

Bartlett, Sir F. C. (1932). *Remembering.* Cambridge University Press.
Gomulicki, B. R. (1952). Recall as an abstractive process. D.Phil. Thesis, University of Oxford.
Kay, H. (1951). Learning of a serial task by different age groups. *Quart. J. Exp. Psychol.* III, 166–83.
Kay, H. (1954). The effects of position in a display upon problem solving. *Quart. J. Exp. Psychol.* VI, 155–169.

(*Manuscript received* 2 *March* 1954)

CHANGES IN MOTOR RESPONSE WITH INCREASE AND DECREASE IN TASK DIFFICULTY IN NORMAL INDUSTRIAL AND PSYCHIATRIC PATIENT SUBJECTS[1]

By PETER H. VENABLES

Medical Research Council, Unit for Research in Occupational Adaptation, Maudsley Hospital

I. *Introduction* (p. 101). II. *Method* (pp. 101–104). III. *Results* (pp. 104–108). IV. *Discussion* (pp. 108–109). V. *Summary* (p. 109). *References* (p. 110).

I. Introduction

Clinical observation led Bleuler (1906) to describe two classes of people. In one, where 'sanguine' temperament is displayed, there is a quick and intensive reaction to emotional impression, in the other the affect is suppressed, leading to displacements and conversions. Nunberg (1918) produced experimental evidence for these two types in work on motor movement in response to word association. The first type made excessive, and the second decreased, motor movement under affect. Berrien (1939) found evidence of two extreme types of reaction to emotion-provoking stimuli in rhythmical finger movement of psychotic patients. He described these as excitatory and inhibitory in nature.

In work on the performance of pilots in a simulated aircraft cockpit, Davis (1948) observed two sorts of errors, those of overaction, and those of inertia. Subjects displaying overactivity reaction made larger control movements and excessive reaction to instrument deviation; the extent and gradient of movement was greatly increased and overcorrecting frequent. Subjects showing inertia reaction made errors which were large and of long duration; control movements were relatively small, and were less disturbed by restless movement than those in the performance of subjects showing overactivity reaction. As a further test of this finding of an 'overactivity-inertia' continuum, Davis devised a less complicated task involving pointer alignment. Similar reactions were observed on this task. The relationship of these reactions to personality type was found to confirm previous work. Anxious subjects were found to show overactive, and hysterics inert reactions.

The present experiment was designed to provide further data bearing on the hypothesis that (*a*) behaviour on a task may vary on a continuum ranging from overactive to inert; and (*b*) that this behaviour is related to personality factors in such a way that when the task is made more difficult dysthymics tend to show modification of performance in the overactive direction, while hysterics change to a more inert form of behaviour. In addition, further evidence was sought on the direction of change of behaviour when task difficulty is decreased.

II. Method

The experiment was conducted as part of a larger investigation into the effects of psychological handicap. It occupied approximately 8 minutes during a session lasting $1\frac{1}{2}$ hours during which the subject completed a battery of intelligence and personality tests. The

[1] Based on part of a thesis approved for the degree of Ph.D. in the University of London.

short duration of the experiment was not deemed to be a disadvantage in that it offered a possibility to test previous workers' findings under conditions not previously employed, and provided a test of the possibility of generalization.

The type of experiment employed was dictated by the practical aim of simulating the stimulus conditions met with by a bus conductor in the course of his job. It provided a means of measuring the subject's reaction to an increase or decrease in the rate of presentation, or in complexity of stimuli. This was an attempt to parallel change in 'load' conditions imposed by an increase or decrease in the number of passengers boarding the bus.

Subjects. Sample 1 consisted of 210 male trainees for the job of bus conductor. Their age range was 18–46 years, mean $28{\cdot}1 \pm 7{\cdot}5$. Data were available for each subject to enable factor scores of Intelligence, Neuroticism, and Introversion-Extraversion to be allocated to him. These scores were derived from the analysis of results on 15 psychometric tests.

Sample 2 had the purpose of providing validation groups to substantiate findings obtained on sample 1. The subjects in sample 2 consisted of 11 showing anxiety states (age range 21–48 years, mean $30{\cdot}7 \pm 7{\cdot}3$) and 11 hysterics (age range 21–46 years, mean $32{\cdot}8 \pm 8{\cdot}5$) selected from a population of 52 male psychoneurotic patients interviewed at a psychiatric hospital. The selection was made by taking those subjects where clear psychiatric diagnoses showing absence of features other than anxiety or hysteria were confirmed by scores obtained on a questionnaire which had been found previously (Hildebrand, 1953) to differentiate anxiety states from hysterics. The two groups are known as Anxious (A) and Hysterics (H) for purposes of reference in the text. To enable results from samples 1 and 2 to be compared, two groups of 11 subjects each from sample 1 were selected on the basis of high scores on the neuroticism factor and extreme scores on the introversion-extraversion factor; these groups, chosen to have characteristics which tended to approach those of the anxious and hysteric hospital patient groups, may be called quasi-anxious (QA) and quasi-hysteric (QH). The terms quasi-hysteric and quasi-anxious have been adopted in view of the possibly misleading connotation of other labels which might have been used, e.g. industrial hysteric. The groups should be considered to be operationally defined, although it is hoped that they have some resemblance to those from sample 2.

Apparatus. From the subjects' viewpoint the apparatus consisted of a milk-white screen $8\frac{3}{4} \times 5\frac{1}{2}$ in., set in a brass frame, at an angle of 45° from the base. On either side of the centre line of the frame and at $2\frac{3}{8}$ in. from it were two marks over which the stimulus lights appeared. In the horizontal shelf in front of the display screen was a slot $8\frac{1}{2}$ in. long through which a knob protruded. A pointer attached to the knob indicated its position against the marks on the display panel. Five stimulus lights were set vertically behind the display panel above each lateral mark. Their position was not visible until alight. Reading from the bottom the stimuli were (1) white, (2) green, (3) red, (4) showing the word YES, (5) showing the word NO.

The subject was given the instructions, 'When you hear this click' (a signal indicating that new stimulus was being presented), 'move the pointer from the mark in the centre to the mark under the white light, if it is on its own, or has green, or YES over it. However, if it has red or NO over it move it to the mark on the opposite side. After this, move the pointer as quickly as possible to the centre mark'.

The pointer was attached to a lever controlling a pen which recorded the subject's movement on a moving paper. The stimuli were automatically presented in predetermined sequence via the contacts of three uniselectors. The presentation of stimuli was divided into three consecutive periods of 50 displays each. In the first easy period *A*, white lights only were used at a 2 sec. periodicity; this was followed by a difficult period *B*, using all the lights as 1·5 sec. per display, and a final easy period *C*, identical with *A*. The whole task thus took 275 sec. to complete. The order of presentation of lights was such that each direction of movement to each type of display was balanced. Every 5th response in the difficult period *B* was to a simple white light stimulus. This scheme was adopted as it was required to measure characteristics of response which were not themselves reactions to difficult stimuli, but were set in an environment of difficult stimuli. By this means it was hoped to measure not the immediate effect of distraction, but the effect on response to a standard stimulus of the emotional condition produced by having to respond to more complex stimuli.

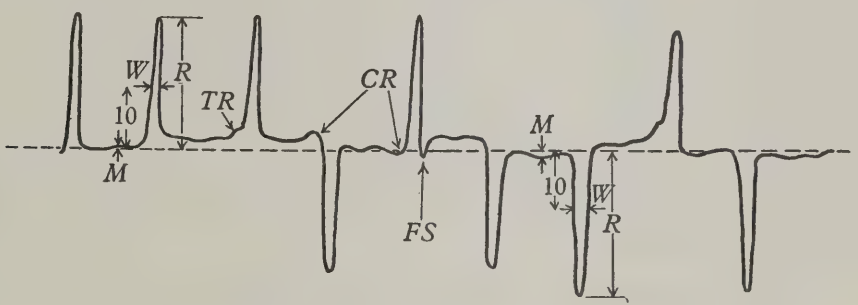

Fig. 1. Artificial example showing method of scoring psychomotor task records.

Procedure. The subject was seated in front of the apparatus which was placed on a low table, so that his forearm was parallel to the ground; he was asked to use his preferred hand. The task was explained to the subject, who was allowed to practise at his own speed whilst 50 displays were presented. It was then explained that his responses would be recorded and that he should expect a period of white lights only at a slow speed, a period using all the lights at a slightly faster speed, and finally white lights only again at a slow speed, there being no break between the periods.

The scoring of the records. Thirty of the 150 responses recorded for each subject were used for measurement. These were every 5th response and were to simple white light stimuli throughout. The method of marking is shown in Fig. 1. The following measures were used and are designated by symbols used throughout the remainder of the paper:

R Maximum extent of response.

W Width of response measured at a fixed distance (10 arbitrary units) from the centre line.

M The deviation from the centre line of the point of return after each response.

Ten responses were measured in each period *A*, *B* and *C*, and means and standard deviations were calculated for each variable. The means were designated $\bar{R}$, $\bar{W}$ and $\bar{M}$ and the standard deviations, giving measures of variability within each period, were denoted by S_R, S_W and S_M.

In addition to this, measurement of selected responses secondary features of all the responses in each period were counted. These were:

FS False starts, where the subject started initially in wrong direction.

TR Trailing returns, where the subject did not make a simple direct return movement to the centre line after a response.

CR Corrected returns, where the return movement overshot the centre mark and had to be corrected.

III. Results

(*a*) *Establishment of a method of measuring the overactive-inert continuum*

Prior to the commencement of any measurement, examination of records of the subjects in sample 1 showed that they ranged from those showing extensive responses of short duration with considerable overshooting and subsequent correction, to those showing short responses lasting some length of time. These were thought to represent opposite poles of the overactive-inert continuum. A period of 25 typical responses (nos. 10–34) in period *A* was rated for this feature on a five-point scale by the writer and an assistant. The correlation between the ratings was 0·807. There was thus a feature of the records which could be determined with reasonable reliability. The problem was to decide how this feature might best be measured so that changes between the different periods of the task could be followed. The possible measures of activity could be derived from those measures already outlined; these were:

(1) The amount of response per unit time, measured by $\bar{R}/\overline{W}$.

(2) The magnitude of the response, measured by $\bar{R} \times \overline{W}$.

The term magnitude of response is somewhat ambiguous and needs further explanation. Measured by $\bar{R} \times \overline{W}$, it is proportional to the area of the response. Thus a response of small magnitude is one of small extent, or one which is completed quickly, or both. The other measure $\bar{R}/\overline{W}$, where the extent of response is divided by the time taken for it, gives a measure which is of course related to 'magnitude' but is not identical. A response showing 'large response per unit time' is one where a response of large extent is completed quickly. We should expect a positive correlation between small magnitude of response and large response per unit time, although in so far as there is a difficulty of choice between the two measures this correlation should depart from unity. The problem of deciding which of these measures was more suitable was tackled in the following way.

It was thought that use could be made of the secondary features of responses along with the rating determined above to provide evidence on which a decision might be made. It was hypothesized that an 'active' subject might show a greater number of false starts (*FS*) and corrected returns (*CR*), and a smaller number of trailing returns (*TR*) than an inert subject. With this in mind $\bar{R}/\overline{W}$, $\bar{R} \times \overline{W}$, *FS*, *TR*, *CR*, and the rating of overactivity-inertia (Rat. O-I), scores for period *A* were intercorrelated. The table of intercorrelations (Table 1) shows that R/W has the highest correlation with the rating, and in addition that the pattern of correlations of R/W with *FS*, *TR* and *CR* follows more closely that of the rating with these variables than does that of $R \times W$. These factors determined the choice of R/W as a suitable measure of 'activity'. To add further confirmation a Lawley (1943)

maximum likelihood type factor analysis was carried out on the intercorrelation between R/W_A, FS_A, TR_A, CR_A and the rating. The result is shown in Table 2.

Table 1. *Intercorrelation of hypothesized measures of overactivity-inertia, sample* 1, $N = 210$

		R/W_A	$R \times W_A$	FS_A	RT_A	CR_A	*Rat. O-I*
Large response per unit time	R/W_A	—	—	—	—	—	—
Small magnitude of response	$R \times W_A$	0·694	—	—	—	—	—
Many false starts	FS_A	0·081	0·500	—	—	—	—
Few trailing returns	TR_A	0·221	−0·078	0·337	—	—	—
Many corrected returns	CR_A	0·491	0·237	0·040	0·129	—	—
Overactivity	*Rat. O-I*	0·858	0·618	0·001	0·379	0·437	—

Note: A correlation of 0·136 is significant at the 0·05 level.

Table 2. *Unrotated factor loadings of hypothesized measures of overactivity-inertia*

		I°	II°
Large response per unit time	R/W_A	0·924	0·269
Many false starts	FS_A	0·089	−0·226
Few trailing returns	TR_A	0·342	−·0231
Many corrected returns	CR_A	0·511	0·023
Overactivity	*Rat. O-I*	0·905	0·184
	% variance accounted for	41·18	4·4

The first unrotated factor accounts for 41·18% of the variance, and is most highly loaded on R/W and the rating. It was not possible to identify the second factor, nor was it thought worth while to extract further factors which were shown to be present in the residual matrix by Lawley's test of significance.

The variable R/W_A resulted from measures taken on responses 1, 6, 11, 16, etc. A measure of reliability can be calculated by comparing this score with a similar one taken from responses 3, 8, 13, ..., etc. When this was done a reliability coefficient of 0·970 was obtained. It is clear from this work that a factor of overactivity-inertia was present in the records and that it could be reliably measured.

It was hoped that a measure of 'disorganization of behaviour' independent of the 'activity' factor described above might be developed. The hypothesis could be advanced that a record showing organization was one where the responses, however large or small, were of even length and width and showing accuracy of alignment to the centre line. Disorganization would then be characterized by high M, S_R, S_W and S_M scores. Following the previous procedure the records were rated on a 5-point scale for 'disorganization of response'. The reliability of the rating was 0·667. Measures S_R, S_W, S_M, M and the rating were intercorrelated, with R/W as a marker for the activity factor. The factor analysis of the matrix yielded a first factor accounting for 23·5% of the variance. This having its highest loading on R/W was again clearly an activity factor, the second factor could not be interpreted in an orthogonal position, and when an oblique solution through the cluster S_R, S_M, and the rating was adopted, the correlation between the factors was approximately 0·87. It was therefore apparent that no measure of disorganization could be developed independently of the 'Activity' factor.

(b) Determination of the relation of the overactive-inertia continuum of motor behaviour to personality classification.

Five types of measures were calculated for each subject for each of the nine variables outlined earlier and for the index R/W. These were:

(1) Performance in period A, e.g. R/W_A.
(2) Performance in period B, e.g. R/W_B.
(3) Performance in period C, e.g. R/W_C.
(4) Change in performance between periods A and B, e.g. R/W_{A-B}.
(5) Change in performance between periods B and C, e.g. R/W_{B-C}.

These measures were correlated with factor scores measuring Intelligence, Neuroticism, and Introversion-Extraversion for the group of 210 normal subjects, sample 1.

It was found that increase in Inertia from period A to period B measured by R/W_{A-B} was positively correlated with the factor of Extraversion at a significant level ($r = +0{\cdot}141$, $P < 0{\cdot}05$), and that increase in inertia between period B and period C (R/W_{B-C}) was negatively correlated with Extraversion ($r = -0{\cdot}147$, $P < 0{\cdot}05$). No significant correlations with the other two personality factors were observed.

In the case of the neurotic group, sample 2, rank correlations between the same variables and the anxious-hysteric dichotomy were calculated. It was found that inertia in period A (R/W_A) was significantly correlated with hysteria ($\tau = +0{\cdot}36$, $P < 0{\cdot}05$) and that increase in inertia from period B to C was negatively correlated with hysteria ($\tau = -0{\cdot}41$, $P < 0{\cdot}05$).

The results from 22 subjects from sample 1 selected to be comparable with the anxious and hysteric groups showed that, as in the parent population, increase in inertia from periods A to B, was correlated with hysteria (neurotic–extraversion) ($\tau = +0{\cdot}57$, $P < 0{\cdot}05$) and increase in inertia from periods B to C was negatively correlated with hysteria, ($\tau = -0{\cdot}63$, $P < 0{\cdot}05$).

The rank correlation method used is that developed by Whitfield (1947) from Kendall's (1938) work, for use where one variable is ranked and the other dichotomous. It should be interpreted in this case in the same manner as a biserial coefficient. The use of this method was dictated by the non-normal distribution resulting from the examination of extreme cases.

Mean scores for R/W, R, and W for periods A, B and C, and for change in performance between periods A and B, and B and C are given in Table 3. These are shown for the total normal sample for groups QA and QH derived from it, and for groups A and H from the psychiatric patient sample, sample 2. Mean scores which are significantly different in groups QA and QH and groups A and H in italics. The significance of the difference between scores on these sub-groups was tested by the median method (Moses, 1952); in view of the non-normal distribution of measures. Fig. 2 shows in graphic form the relationships outlined above and shown in the upper part of Table 3.

If we examine the results from either the whole of sample 1 or the subgroups drawn from it, it is seen that extravert and quasi-hysteric subjects show a significant tendency to change their performance in the inert direction under increased difficulty. The reverse tendency is shown in introvert, quasi-anxious, and anxious subjects The hospitalized, hysteric patient presents an initially inert type of response. Commencing performance on this level he does not change his performance in the inert direction under increased difficulty, but maintains a performance which is substantially similar to his starting performance.

Table 3. *Mean performance scores of subjects from samples* 1 *and* 2 *on variables R/W, R, and W, for periods A, B, and C*

Variable	Sample 1			Sample 2	
	Total	QA	QH	A	H
R/W_A	5·93	5·32	6·05	*6·09*	*4·33*
R/W_B	5·38	5·71	5·59	7·03	4·39
R/W_C	5·58	4·82	6·00	6·08	4·72
R/W_{A-B}	+0·55	*−0·39*	*+0·46*	−0·94	−0·06
R/W_{B-C}	−0·20	*+0·89*	*−0·41*	*+0·95*	*−0·33*
R_A	21·37	20·97	21·68	*20·66*	*19·35*
R_B	20·62	20·50	20·73	19·84	19·10
R_C	21·40	21·38	21·78	20·55	19·76
R_{A-B}	+0·75	+0·47	+0·95	+0·82	+0·25
R_{B-C}	−0·78	−0·88	−1·05	−0·71	−0·66
W_A	3·93	4·33	3·76	4·30	4·95
W_B	4·18	3·83	3·92	*3·76*	*4·83*
W_C	4·17	4·66	3·82	4·56	4·57
W_{A-B}	−0·25	*+0·50*	*−0·16*	+0·54	+0·12
W_{B-C}	+0·01	*−0·83*	*−0·10*	*−0·80*	*+0·26*
N	210	11	11	11	11

Differences significant at less than the 0·05 level of confidence are in italic.

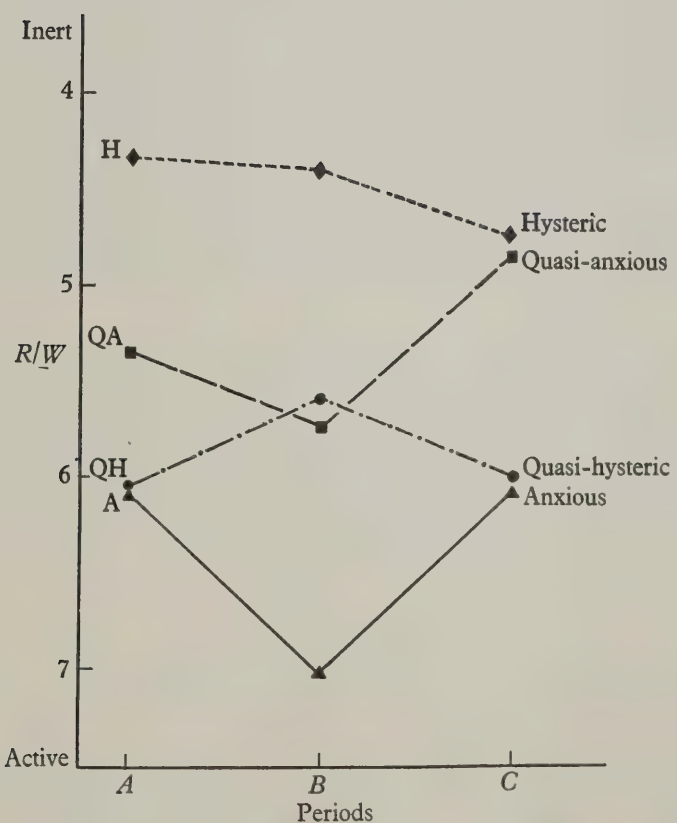

Fig. 2. Changes in activity of response (R/W) between periods A, B, and C in groups A, H, QA and QH.

With decrease in task difficulty there is in all cases a significant tendency for the extravert or hysteric subject to perform more actively whilst the introverted or anxious patient subject shows a decrease in activity.

On examination of the measures which make up the index R/W it is seen that response length tends to shorten in the difficult period, and lengthen subsequently, although to different extents, in different groups of subjects. On decrease of difficulty there is a general lengthening of response in all groups. It is the measure W, the time for which the response lasts, that seems to be responsible to a greater extent for differential behaviour between the groups. In both dysthymic groups there is a speed up of response under difficulty followed by a slowing down with decrease in difficulty. In the hysteric patient group performance is slow initially whilst in the extraverted group drawn from sample 1 there is a slowing down on increase of difficulty. In both of these groups there is a speed up on decrease of difficulty.

While W, measuring speed of response, appears to parallel many of the findings of R/W in dysthymic-hysteric group differentiation it is felt that the latter measure does operate sufficiently differently to justify the reasons for its original inclusion. Evidence is given when motor task variables are related to G.S.R. measures. The work of White (1930) Freeman & Simpson (1938) and Staudt & Kubis (1948) shows that increase in G.S.R. conductance accompanies increase in muscular tension. Other work reported elsewhere (Venables, 1955), indicates that increase in 'activity' measured by R/W_{A-B} is significantly correlated ($P = 0{\cdot}045$ and $0{\cdot}03$) with increase in conductance, in both normal and neurotic groups. Increase in speed measured by W_{A-B} does not show significant correlations with increase in conductance. The work of Freeman & Pathman (1942) and Champion (1950) would lead us to expect that an increase in motor movement would result in a dissipation of autonomic disturbance. In the case of G.S.R. response after the difficult period, it is the measure W_{B-C} and not R/W_{B-C} which shows a significant correlation with decrease in conductance supporting Freeman & Pathman's and Champion's work. It would thus seem that there is a useful differentiation between the measures R/W and W, and that the former may measure behaviourally something akin to muscular tension, whilst the latter measures speed of response.

Correlations of task variables with Intelligence and Neuroticism show that they are in general independent of these factors. If a group of subjects from sample 1 having ambivert and low neuroticism factor scores are compared with the two patient groups the finding of independence of the overactivity-inert continuum from neuroticism is supported by insignificant rank correlations of $+0{\cdot}05$, $+0{\cdot}18$ and $0{\cdot}00$ with R/W_A, R/W_{A-B}, and R/W_{B-C} respectively. Independence from age is shown by correlations of $+0{\cdot}09$, $+0{\cdot}01$ and $+0{\cdot}01$ with the same variables.

IV. Discussion

In the work described an attempt has been made to provide further data bearing on the relation between motor performance and personality. The work of previous writers has been largely confirmed, and additional findings showing evidence of regulatory behaviour after task difficulty has ceased have been given. It is of some interest that in many ways it is the change in performance when the stimulus situation changes from difficult to easy, which seems to be the most valuable indicator of different functions in different types of

personality. Whilst in the hysteric group there was little change when the task difficulty increased, it was the marked change in the direction of overactivity on decrease of difficulty which differentiated this group from the anxious patients.

An external source of evidence on this point is given when the performance scores of 18 subjects in sample 1 who were discharged (largely for dishonesty) are examined. The scores for these subjects are compared with those resigning or remaining in employment in the same period of six months. It was found that the measure R/W_{B-C} differentiated the discharged from the rest at a significant level ($P < 0{\cdot}01$). This was not, however, the case with R/W_{A-B} where a quite insignificant F-ratio was obtained. The discharges show behaviour which was in many ways similar to that of extraverted subjects.

The results of this work have some similarities to those of Foulds (1952) although he used a very different type of experiment. Foulds found that distraction produced a speed up of performance of dysthymics on the Porteous Maze, whilst speed up, although shown in non-dysthymics, was not significant. Examination of Table 3 shows a very similar state of affairs in the performance of our anxious and hysteric groups on periods A and B.

Also shown in the data presented is some indication of the difference between the performance of hospitalized and non-hospitalized groups of tentatively similar personality type. This is particularly striking in the case of the hysterics and quasi-hysterics, where the initial performance of the hysteric in hospital is so inert that no change in the inert direction seems possible. On the other hand, the quasi-hysteric in industry presents an initially quite active performance. These differences should not be pursued too far in view of the present unsatisfactoriness of tests with which it is hoped to measure in a normal population the counterpart of anxiety-hysteria in a hospital population.

The findings quoted indicate that performance is by no means a static phenomenon and that perhaps greater attention should be given in industrial situations to the maintenance of an even level of difficulty of work, in order to avoid alteration of performance due to interaction effects with features of the operatives' 'personality' in its narrowest sense.

Any attempt to present a theoretical basis for these findings is at the moment considered premature; further experimentation to test various hypotheses is proceeding.

V. Summary

A motor task, involving the movement of a pointer in relation to stimulus lights under easy, difficult and subsequently easy conditions was performed by 210 male subjects consisting of a random selection of trainee bus conductors, age range 18–46, and by 22 male patients, age range 21–48, in a psychiatric hospital, selected from a group of 52 patients as having clear-cut psychiatric and questionnaire diagnosis.

Performance on the motor task could be described as falling on a continuum of overactivity-inertia. In general, under conditions of increased task difficulty, extraverted subjects showed a tendency to increased inertia of performance and introverted subjects to increased activity. When the task became easier the reverse relationship was found to hold true.

The author wishes to acknowledge the valuable assistance received from Dr H. J. Eysenck and Mr M. B. Shapiro and his colleagues Drs Eisler, Heron, Markowe, O'Connor and Tizard. The psychiatric patient subjects were made available through the co-operation of Dr M. Desai.

References

Berrien, F. K. (1939). Finger oscillations as indices of emotion. I. Preliminary validation. *J. Exp. Psychol.* xxiv, 485–98.

Bleuler, E. (1906). *Affektivität, Suggestibilität, Paranoia.* Halle: Verlag Carl Marhold.

Champion, R. A. (1950). Studies of experimentally induced disturbance. *Aust. J. Psychol.* ii, 90–9.

Davis, D. R. (1948). *Pilot Error—Some Laboratory Experiments.* London: H.M.S.O.

Freeman, G. L. & Pathman, J. H. (1942). The relation of overt muscular discharge to physiological recovery from experimentally induced displacement. *J. Exp. Psychol.* xxx, 161–74.

Freeman, G. L. & Simpson, R. M. (1938). The effects of experimentally induced muscular tension upon palmar skin resistance. *J. Gen. Psychol.* xviii, 319–26.

Foulds, G. A. (1952). Temperamental differences in maze performance. Part II. The effect of distraction and of electroconvulsive therapy on psychomotor retardation. *Brit. J. Psychol.* xliii, 33–41.

Hildebrand, H. P. (1953). A factorial study of introversion-extraversion by means of objective tests. Ph.D. Thesis, University of London.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, xxx, 81–93.

Lawley, D. N. (1943). The application of the Maximum Likelihood Method of Factor Analysis. *Brit. J. Psychol.* xxxiii, 172–5.

Moses, L. E. (1952). Non-parametric statistics for psychological research. *Psychol. Bull.* xlix, 122–43.

Nunberg, H. (1918). On the physical accompaniments of association processes. In Jung, C. G., *Studies in Word Association.* London: Heinemann.

Staudt, V. & Kubis, J. F. (1948). The psychogalvanic response and its relation to changes in tension and relaxation. *J. Psychol.* xxv, 443–53.

Venables, P. H. (1955). Some findings on the relationship between g.s.r. and motor task variables (To appear in *J. Gen. Psychol.*).

White, M. M. (1930). Relation of bodily tension to electrical resistance. *J. Exp. Psychol.* xiii, 267–77.

Whitfield, J. W. (1947). Rank correlation between two variables one of which is ranked, the other dichotomous. *Biometrika*, xxxiv, 292–6.

(*Manuscript received* 21 *April* 1954)

THE IDEA OF A DISTRIBUTION

By JOHN COHEN and C. E. M. HANSEL
Department of Psychology, University of Manchester

I. *Introduction* (p. 111). II. *Method* (pp. 111–112). III. *Type I: procedure and results* (pp. 112–115). IV. *Type II: procedure and results* (p. 116). V. *Type III: procedure and results* (pp. 116–120). VI. *Discussion of results* (pp. 120–121).

I. Introduction

The experiments to be described form part of the study of what we propose to call 'subjective probability'. In everyday life we are constantly choosing, estimating, predicting, judging, or deciding on the basis of incomplete knowledge. In all these activities there is an element of uncertainty on the part of the individual. In so far as the individual prefers one alternative to the other or considers it more likely to a greater or less extent, his preferences may perhaps be evaluated. We shall inquire whether judgements based on partial information follow their own psychological rules in more or less consistent fashion, and how these rules compare with rules of mathematical probability based on logical criteria. A second aim of this inquiry is to trace the systematic or other changes in the forms of subjective probability characteristic of different ages during the course of development. It is of particular interest to examine the way the child assimilates statistical rules in so far as they are repeatedly verified in experience from the earliest days of life. In this first series of experiments we shall confine ourselves to the age range 10 to 15 years. We should add at this point that we are here concerned with the judgement as such. In further elaboration of these studies it will be important to distinguish between the judgement itself and the intensity of conviction with which it is held.

II. Method

In the present paper we examine the types of 'guess' which are made in situations of increasing complexity. These range from choices between two alternatives to judgements which may implicitly involve ideas of a frequency distribution. We also refer to conditions in which the amount of information provided is systematically controlled. In one experiment, for example, we show a group of subjects a large urn containing an unspecified number of blue or yellow beads. They are not told the proportions of blue and yellow beads in the urn. The beads are thoroughly mixed and we take out 4 and place them in a beaker. This is repeated until there are 16 beakers each containing 4 beads. The subjects might then be asked how many of the beakers contain, say, exactly 2 blue beads. From the data given by the subjects, the distribution of beakers containing 0, 1, 2, 3, and 4 blue (or yellow) beads can be built up and compared with a theoretical distribution.

The particular distributions to be studied in these experiments may be built up from subjects' estimates in three different ways. These vary with the kind of judgement made by the subject. The three different ways are:

Type I. The beads are withdrawn from the urn in sets of 4. They are placed individually in separate beakers and the subject guesses the colour of each bead.

Type II. The beads are withdrawn from the urn in sets of 4 and each set of 4 is placed

in a separate beaker. There are 16 beakers in all. The subject guesses the number of blue and yellow beads in each beaker.

Type III. The beads are placed in 16 separate beakers, as in II above. The subject now guesses the number of *beakers* containing 0, 1, 2, 3 and 4 beads respectively.

In the present series of experiments, which are analysed in detail, the proportions of blue and yellow beads in the urn are stated to the subjects to be unknown. We refer briefly to other experiments in which the subjects are told that definite proportions of beads are in the urn.

The total number of subjects studied in these experiments is about 330. About 100 subjects took part in experiments relating to Types I and II and about 230 in experiments relating to Type III.

III. Type I: procedure and results

The procedure takes the form of a class demonstration in which beads are drawn singly from a large urn containing blue and yellow beads in unspecified numbers, and the subjects guess the colour of each bead.

We say to the subjects:

> There are a lot of beads in this urn but we do not know exactly how many. They may be blue or yellow, but no other colour. There may be no blue beads or any number of blue beads. There may be no yellow beads or any number of yellow beads. We do not know the proportions of yellow and blue beads in the urn. We shall take 4 beads at a time from the urn and place each one of these in a separate beaker. There are 4 beakers here, one for each bead.

The subjects can hear the beads drop into the beakers but they cannot see the colour of the beads at any stage of the experiment. They record their guess of the colour of the bead in each of the four beakers. The beads are then replaced in the urn. This procedure is repeated 16 times.

The 95 subjects yield 1520 arrangements of a set of 4 beads. In Table 1 we show in col. (i) the 16 possible arrangements of the 4 beads. Col. (ii) gives the number of times each of the 16 arrangements occurs in the choices made by the group. Col. (iii) gives the theoretical number of times each arrangement would occur if the ratio of blue to yellow beads were based on the respective number allotted by the group (3162 blue and 2918 yellow) and not on a 1:1 ratio. On the assumption of a 1:1 ratio the theoretical number of times each arrangement would occur would invariably be 95. The values in col. (iii) may not be directly comparable with those in col. (ii) because individual subjects may be basing their choices on implicit assumptions about differing proportions of blue and yellow beads in the urn. In so far as such implicit assumptions are made they may fluctuate from time to time during the course of the experiment. The effect of these varying assumptions would be to give a flattened theoretical distribution as compared with the distribution based on group values. In view of these considerations we show in col. (iv) the theoretical number of times each of the 16 arrangements will occur if the ratio of blue to yellow beads is based on the number of each colour allocated by each individual rather than on the totals of the group as a whole. The values in col. (iv) were obtained by calculating a binomial distribution for each individual in terms of the proportions of blue and yellow beads derived from his total choices, then summing the results for all 95 subjects.

To return for a moment to col. (iii) in Table 1, we should note that the proportions of blue and yellow beads in the urn were stated by us to be unknown. We find that the

proportions allotted by the subjects tended on the whole towards 0·5, the actual values for the group being 0·52 (blue) and 0·48 (yellow) with a standard deviation of 0·06. We should expect that unless any special influence affected the number of choices of one rather than the other colour, the proportions implicitly allotted to them would tend to 0·5. In fact the proportions 0·52 and 0·48 depart significantly from this mean value. From the observed number of choices, 3162 blue and 2918 yellow, χ^2, calculated from the expected number, 3040, is 9·8 ($P < 0·01$). This may be due to a number of causes such as colour preference, a preference for writing *B* rather than *Y*, or to the wording of the instructions.

Table 1. *Choice of all possible arrangements of 4 beads, each of which may be blue or yellow*

(N = 95. Mean age 10 years 6 months.)

	Frequencies		
Arrangements (i)	Observed (ii)	Calculated from proportions of colours allotted by the group* (iii)	Calculated from proportions of colours allotted by each individual (iv)
B B B B	151	111·1	123·8
B B B Y	116	102·6	102·9
B B Y B	83	102·6	102·9
B Y B B	77	102·6	102·9
Y B B B	99	102·6	102·9
B B Y Y	102	94·7	92·0
B Y B Y	93	94·7	92·0
B Y Y B	88	94·7	92·0
Y B B Y	123	94·7	92·0
Y B Y B	75	94·7	92·0
Y Y B B	76	94·7	92·0
B Y Y Y	74	87·4	86·7
Y B Y Y	62	87·4	86·7
Y Y B Y	90	87·4	86·7
Y Y Y B	93	87·4	86·7
Y Y Y Y	118	80·7	85·6

* These expected values are calculated from a population of blue and yellow beads in the proportions 0·52 and 0·48 respectively, representing the colour preferences of the group.

If we now compare the observed values in col. (ii) with those shown in col. (iv), calculated on the basis of proportions of colours allotted by each individual, we find that they differ significantly. χ^2 is 57·25, D.F. = 15, $P < 0·001$. This means that even when we take the individual allotments of colour fully into account the distribution of arrangements is not random. Other factors must therefore be operative. Inspection of the values in col. (ii) suggests that one such factor may be a preference for particular kinds of arrangement.

We can examine this suggestion in the first instance by comparing the preference for different arrangements within each of the five subgroups, each subgroup containing 0, 1, 2, 3 or 4 blue beads, disregarding order. Within any subgroup, if there were no preference for any particular arrangement, the expected number for each arrangement would be the same. We calculate the expected values from the total number of choices within the subgroup divided by the number of possible arrangements in the particular subgroup (see Table 2). $\chi^2 = 35·1$, D.F. = 15, $P < 0·01$, thus indicating a preference for particular arrangements.

The data indicate thus far that choices are not made at random but vary in a systematic manner within each subgroup where the number of beads of each colour remains constant. We may now ask what it is that determines this preference for particular kinds of arrangement.

Table 2. *Choice of arrangements of* 0, 1, 2, 3 *or* 4 *blue beads disregarding order*

(N = 95 Mean age 10 years 6 months.)

Arrangements (i)	Frequencies: Observed (ii)		Frequencies: Calculated by equal division of subgroup total (iii)
B B B B	151	151	151·00
B B B Y	116		93·75
B B Y B	83	375	93·75
B Y B B	77		93·75
Y B B B	99		93·75
B B Y Y	102		92·83
B Y B Y	93		92·83
B Y Y B	88		92·83
Y B B Y	123	557	92·83
Y B Y B	75		92·83
Y Y B B	76		92·83
B Y Y Y	74		79·75
Y B Y Y	62		79·75
Y Y B Y	90	319	79·75
Y Y Y B	93		79·75
Y Y Y Y	118	118	118·00
Total	1520	1520	1520·00

One possible factor affecting the preference for certain arrangements is their balanced or symmetrical character. From the point of view of symmetry we distinguish three kinds of arrangement:

(i) B B B B or Y Y Y Y; B Y Y B or Y B B Y } balanced

(ii) B Y B Y or Y B Y B; B B Y Y or Y Y B B } alternating or repetitive

(iii) B B B Y or Y Y Y B; B B Y B or Y Y B Y; B Y B B or Y B Y Y; Y B B B or B Y Y Y } asymmetrical

When the observed values for the three groups are separately compared with the corresponding frequencies in col. (iv) of Table 1, we find that the observed value of the first group exceeds its theoretical ($\chi^2 = 19$, D.F. $= 1$, $P < 0{\cdot}001$); the observed value of the second is not appreciably different from its theoretical; the observed value of the third group is significantly lower than the theoretical ($\chi^2 = 5{\cdot}46$, D.F. $= 1$, $P < 0{\cdot}02$). This seems to indicate a preference for the balanced and an avoidance of the asymmetrical kind of arrangement.

Another factor might be the tendency in any given arrangement to repeat the previous choice of blue or yellow rather than to change to the other colour. The total number of

choices, 1520, yields 4560 possible changes, i.e. three for every arrangement. There are in fact 2110 choices in which there is a change of colour and 2450 choices in which the colour is not changed. This difference is significant ($\chi^2 = 27{\cdot}1$, D.F. $= 1$, $P < 0{\cdot}0001$) and indicates a preference for choosing the same colour after any given choice within a set of 4 beads.

Although the subjects make their choices in relation to individual beads in sets of 4, we can combine the sets in such a way as to yield a distribution of arrangements containing 0, 1, 2, 3 or 4 blue beads. The subjects' choices classified in this way are shown in Table 3, together with the corresponding values calculated on the basis of the proportions of beads of each colour allotted by each subject.

Table 3. *Distribution of arrangements containing* 0, 1, 2, 3 *or* 4 *blue beads*

	Number of blue beads					
	0	1	2	3	4	Total
Observed values						
Total	118	319	557	375	151	1520
Mean	1·2	3·4	5·9	3·9	1·6	16
Calculated values*						
Total	85·6	346·8	552·0	411·6	123·8	1520
Mean	0·9	3·7	5·8	4·3	1·3	16

* Based on proportions of blue and yellow allocated by each subject.

If the subjects were making their choices at random, even if their preferences for the two colours were uneven, the two distributions in Table 3 would approximate to each other. The fact that the distributions differ significantly implies that the subjects are not behaving in random fashion. We have already suggested two factors that might account for this.

The comparison between the two distributions in Table 3 is given not to suggest that the children in this experiment had any notion of a binomial distribution, but to demonstrate that they were not merely allocating colours at random. Any resemblance between the two distributions may be due to the fact that apart from the factors isolated there is a very large and diverse number of influences affecting the choices of different individuals about which we know little. The resulting distributions accordingly tend in some measure to have the form of a distribution resulting from random events.

In the foregoing experiment the children guessed the colour of single beads. We might have expected that their choices would have been random and hence the distributions derived from the choices would tend towards the binomial. This does not appear to be the case. On the contrary we find:

(i) The proportions of blue and yellow beads implicit in the choices although varying to some extent have a relatively small scatter. Two-thirds of the proportions (for blue) are within the range 0·46 to 0·58.

(ii) Their choices are not made at random.

(iii) The non-random character of the individual choices is due to preferences for certain types of arrangement within the sets of 4 beads.

(iv) There is a preference for balanced and an avoidance of asymmetrical arrangements.

(v) There is a preference for repeating the choice of any one of two colours within any set of choices.

IV. Type II: procedure and results

We now turn to the second way in which the distributions may be built up. Four beads at a time are placed in turn in each of the 16 beakers and, as they are heard to drop in the beaker, the subjects record their answer to the question (Exp. II*a*): how many blue beads are in the beaker? In Exp. II*b* the subjects were asked to record how many blue *and* yellow (or, reversing the order of the question, yellow *and* blue) beads they thought were in each beaker. The order II*a*, II*b* was followed with one group of subjects and II*b*, II*a* with a second, and the results for each experiment combined.

The distribution of arrangements yielded by the choices is shown in Table 4, together with the calculated values based on proportions of blue and yellow implicitly allotted by each individual.

Table 4. *Observed and calculated distributions*

(N = 93. Age 10 + years.)

Experiment		Number of blue beads: 0	1	2	3	4	Total
II*a*	Observed	92	334	436	363	263	1488
	Calculated	65	283	518	459	163	1488
II*b*	Observed	245	318	361	310	254	1488
	Calculated	117	366	513	365	125	1488

Unlike the Type I experiment the subjects in Type II actually state a distribution for each one of the 16 beakers. Random behaviour on their part would yield an equal number of beakers for 0, 1, 2, 3 and 4 beads. This does not appear, for both Exps. II*a* and II*b* differ significantly from a rectangular distribution. Indeed they show a marked tendency in the direction of a binomial distribution. The subjects' choices imply that they regard 2B and 2Y as more likely than 1 B and 3 Y (or vice versa) or 0 B and 4 Y (and vice versa). Similarly, they regard 1 B and 3 Y or 1 Y and 3 B as more likely than 0 B and 4 Y (or 0 Y and 4 B). In this experiment random choices would lead away from a binomial towards a rectangular distribution. In the previous experiment random behaviour would lead away from a rectangular towards a binomial distribution of arrangements.

One other feature of this experiment should be noted, i.e. the difference between the distributions yielded by Exps. II*a* and II*b* respectively. This may be attributed to the form of the question. It appears that when we ask: 'How many blue beads are in each beaker?' and do not make any reference to yellow beads, the subjects tend to avoid the value 0 blue beads.

We may conclude from this experiment that subjects do not allocate proportions of blue and yellow beads within sets of 4 at random. They exhibit a significant preference for the value 2 B, 2 Y and also for 1 B, 3 Y or 3 B, 1 Y as compared with 0 B, 4 Y or 0 Y, 4 B.

V. Type III: procedure and results

The third way in which a distribution may be built up is by placing 4 beads drawn at random into each one of 16 beakers. In this experiment the subjects are told that the proportions of blue and yellow beads in the urn are unknown. They now guess the number of beakers containing 0, 1, 2, 3 or 4 blue (or yellow) beads, and these guesses constitute their distributions. In one group (aged 11–13) the subjects, after giving their guesses, explained in writing why they chose their particular values.

In previous studies we have found that the order in which the questions are asked in this type of experiment may affect the subjects' estimates, the earlier questions tending to receive higher values than they would otherwise get. The estimates may also be affected by the order in which the colours are referred to in the questions. In order to eliminate these influences it was necessary to employ ten ways in which the questions were presented. In five of these blue was referred to first in the questions and in five yellow was referred to first. In the set of five presentations for each colour the questions were so arranged that all five questions appeared in each position an equal number of times. The ten orders of presentation were given to approximately equal groups of subjects, each subject receiving a single order of presentation. The means for the entire samples were obtained from the subgroup means.

A specimen set of questions is set out below:

NAME: ..... Age: .....

How many beakers will have 4 yellow beads and no blue beads? .....

How many beakers will have no yellow beads and 4 blue beads? .....

How many beakers will have 3 yellow beads and 1 blue bead? .....

How many beakers will have 1 yellow bead and 3 blue beads? .....

How many beakers will have 2 yellow beads and 2 blue beads? .....

Total 16

The mean distributions for the different age groups are given in Table 5.

Table 5. *Beakers containing* 0, 1, 2, 3 *or* 4 *blue beads*

(Mean values allotted as proportions of 16. Unknown proportions of blue and yellow beads in the urn.)

		Number of blue beads					
Age	N	0	1	2	3	4	Total number of beakers
10+	87*	3·2	3·1	3·7	2·9	3·2	16
12+	31†	3·3	2·7	3·5	3·2	3·1	16
14+	26†	3·0	3·0	3·7	3·2	3·2	16

* Including some 11-year-olds. † Girls only.

We wish to see first whether the subjects, when the proportions are unknown, estimate the number of beakers containing 0, 1, 2, 3 or 4 blue beads respectively in a manner that is likely to arise in practice on any assumption that might be made about the contents of the urn. If, for example, a subject estimated that each one of the 16 beakers contained 4 blue beads, that estimate would accord with the assumption that all the beads in the urn were blue. On the other hand, if a subject stated that half of the beakers did not contain a single blue bead and half contained 4 blue beads each, this would be a most unlikely outcome on any conceivable assumption about the contents of the urn: one subject gave these values. We do not know whether the subjects make any explicit assumptions about the contents of the urn, but we may perhaps infer, from the number of blue and yellow beads indirectly assigned by each subject, that they do tend to assume approximately equal proportions. Nevertheless, few of the individual distributions have a high probability of arising in terms of the number of blue and yellow beads implicitly assigned.

In order to assess the effect on the distributions given of increasing the amount of information about the contents of the urn, we have carried out an experiment with a

comparable age group, 10–12 years ($N = 87$), in which the subjects are told that there are equal numbers of blue and yellow beads in the urn. The resulting group distributions are much the same as those obtained here. We have also studied the case in which the subjects are told that the proportions are 3:1 and there the distributions are very different, especially as given by the 14+ group. A detailed analysis of these two experiments will be given in a separate paper.

Secondly, we can inquire whether the subjects at the ages 10+, 12+ and 14+ have any preference for assigning equal values to 0 B and 4 B and/or equal values to 1 B and 3 B, by comparing the actual values assigned to these categories with the values assigned to the remaining categories, taking them in pairs. This comparison is shown in Table 6.

Table 6. *Categories of beakers containing* 0, 1, 2, 3 *or* 4 *blue beads compared in pairs of category*

(N = 87 (10+ years) + 31 (12+ years) + 26 (14+ years).)

Paired categories		Proportions of samples assigning equal values to pairs of categories of beaker			
		10+	12+	14+	Total
0	1	0·17	0·16	0·12	0·16
0	2	0·17	0·13	0·08	0·15
0	3	0·22	0·10	0·19	0·19
0	4	0·16	0·42	0·42	0·26
1	2	0·15	0·23	0·23	0·18
1	3	0·17	0·45	0·35	0·26
1	4	0·26	0·10	0·12	0·20
2	3	0·15	0·16	0·23	0·17
2	4	0·25	0·13	0·12	0·20
3	4	0·15	0·13	0·12	0·14

We can calculate χ^2 from the observed number of instances in which equal and unequal values are assigned by the 12+ and 14+ groups to 0, 4 and/or 1, 3 as compared with the number of equal and unequal values assigned to the remaining pairs. $\chi^2 = 37{\cdot}4$, D.F. $= 1$, $P < 0{\cdot}001$. This tendency to assign equal values to 0, 4 and to 1, 3 does not appear in the 10+ group (see Table 6).

The greater tendency of the older children (12+ and 14+) as compared with the younger ones (10+), to assign equal values to 0, 4 and 1, 3 is indicated by the significance of the χ^2 calculated from the values assigned by the respective groups. This calculation is made by comparing the number of times the different age groups allot equal values to the 0, 4 and 1, 3 categories as compared with all other possible pairs of categories. $\chi^2 = 18{\cdot}7$, D.F. $= 1$, $P < 0{\cdot}001$.

This effect is brought out clearly in the protocols of three subjects of a subsample of 30:

(1) Age 12 years 11 months:

The beakers with one yellow, and 3 blue, and those with 3 yellow and one blue, have opposite amounts so I've made them the same. And I've done the same as that with the ones with 4 blue and no yellow; and 4 yellow and no blue.

	No. of blue beads				
	0	1	2	3	4
Distribution given	4	2	4	2	4

(2) Age 12 years 7 months:

There will be most in number V (2B, 2Y) as there is an even chance for each two. Numbers I (4B, 0Y) and II (0B, 4Y) rule each other out and so I have put them even. Numbers III (3B, 1Y) and IV (1B, 3Y) will rule each other out also. As there is not an even number left I simply put 4 for number III (3B, 1Y) and 3 for number IV (1B, 3Y).[1]

	No. of blue beads				
	0	1	2	3	4
Distribution given	2	3	5	4	2

(3) Age 12 years 10 months:

I thought that the beakers having an equal number of beads in them would not be many and I only put one. Also the first and second answers (0B, 4Y and 4B, 0Y) are likely to be about the same, and the third and fourth answers (1B, 3Y and 3B, 1Y).

	No. of blue beads				
	0	1	2	3	4
Distribution given	4	3	1	2	6

Thirdly, bearing in mind that the subjects implicitly allot approximately equal proportions to blue and yellow, it is of interest to see whether they tend to assign higher values to 2B, 2Y as compared with any other combination. If so, it might indicate that the samples in the beakers are regarded as miniatures of the contents of the urn and hence be allotted equal numbers of blue and yellow beads. This is supported by the fact that one subject assigned all 16 beakers to the value 2B, 2Y. Furthermore, there is a tendency, when the proportions in the urn are specified as 3:1, to allot high values to 3B, 1Y.

We can calculate the values assigned by all three age groups to 2B, 2Y as compared with the values assigned to 0B, 4Y; 1B, 3Y; 3B, 1Y and 4B, 0Y (see Table 7). There is no difference between the age groups in this respect, but 2B, 2Y receives the highest value in a significantly large number of instances. $\chi^2 = 9{\cdot}3$, D.F. $= 1$, $P < 0{\cdot}01$.

Table 7. *Number of times highest value is assigned to the different categories of beaker*

	No. of blue beads				
Age	0	1	2	3	4
10+	8	10	18	9	7
12+	3	1	7	5	5
14+	2	1	5	2	4
Total	13	12	30	16	16

Fourthly, we can see whether there is any significant tendency to assign higher values to 1 as compared with 0 and to 3 as compared with 4. If such higher values were found to be present they might indicate a tendency towards a distribution.

χ^2 calculated from the combined values assigned by all three groups to $1 \gtrless 0$ and to $3 \gtrless 4$ is 9·6, D.F. $= 1$, $P < 0{\cdot}01$. Inspection of Table 8 shows that in all three groups there is a tendency to assign greater values to 0 than to 1, and to 4 than to 3. Instead of

[1] This subject herself numbered the questions I to V.

a tendency towards a distribution, we find a tendency away from it. This might be accounted for by the marked preference for balance and the avoidance of asymmetrical arrangements as demonstrated above in relation to Type I.

Table 8. *Comparison between values assigned to categories of beaker containing* 0 *blue and* 1 *blue, and* 3 *blue and* 4 *blue*

	1, 0			3, 4		
Age	>	=	<	>	=	<
10+	32	15	42	27	16	44
12+	8	6	17	10	6	15
14+	10	3	13	9	3	14
Total	50	24	72	46	25	73

VI. Discussion of results

Perhaps we should first note the effects on subjects' implied assumptions about the contents of the urn of not specifying the proportions. In all three types of experiments the behaviour of the subjects does not, on the whole, display a great deal of scatter around mean values of 0·5 blue and 0·5 yellow. Although they have an exceedingly wide choice of possible values, some 10% of the subjects chose equal values of blue and yellow beads when asked to state what they thought to be the contents of the urn. Their implicit estimates of the contents of the urn, in all three experiments, as derived from their guesses or distributions, also show this limited range of choice. These implicit estimates when the proportions in the urn are not specified resemble the corresponding estimates made in other experiments by comparable groups of subjects when the proportions in the urn are specified as 1:1. When, however, the proportions are specified as 3:1 the implicit estimates are quite different. We may thus conclude that when the proportions are unspecified there is a tendency for each child to assume that the proportions are equal or nearly so. The assumption of equality, under the conditions described, seems to be the simplest and the most naturally acceptable to the child.

In the three types of experiment the subjects' choices, at least the majority of them, are not such as could have occurred by random selection. In Type I there is the preference for certain kinds of arrangement. In Type II the subjects do not tend to allot similar proportions of blue and yellow beads to the different beakers. In Type III the choice of numbers which add to 16 is not random and there is a preference for 2, 2 and for 0, 4.

So far we have been discussing the group means. Some light is thrown upon the individual choices by the explanations given by a group of 12+ ($N=30$), although we do not suppose that these explanations fully account for the choices actually given. Some 20% of the group appear to have some idea of a distribution. They mention in their explanations one or more of the reasons for the various effects which have emerged in the analysis. Some 40% produce some definite reason for the manner in which they have made their choices, such as, for example: 'I wanted to have every number different' or 'I did not think any of the numbers would be alike'. The remaining 40% gave no definite reason for their choices, e.g. 'They were the first numbers which came into my head and added up to 16. When I had done the first three I divided what was left into two as I thought best.'

Even at the age 10+ there is a tendency to assign the highest value to 2B 2Y. But the preference for assigning equal values to 0, 4 and to 1, 3 appears only in the age group

12+ and upwards. It does not appear in the 10+ group. This preference may be an important new step beyond the level reached by the 10-year-olds for it enables them to construct a symmetrical distribution. They seem to be applying a 'principle of indifference' in that they assign equal likelihood to 1 B, 3 Y as to 3 B, 1 Y.

There thus appears to emerge, within the age range 10–15 years, four stages in the development of the idea of a distribution. First, there is the belief that the numbers in the distribution will tend to vary. Secondly, there is the general tendency to assume that the category of equal proportions (2 B, 2 Y) will occur most often, although a few children believe that it will occur least often. Thirdly, there is the tendency to assign the same likelihood to categories with similar structure (1 B, 3 Y = 3 B, 1 Y; 0 B, 4 Y = 4 B, 0 Y). A further stage, which is apparently not yet reached at age 14+ when the proportions are unspecified, is to assign greater likelihood of occurrence to 1 B, 3 Y than to 0 B, 4 Y and to 3 B, 1 Y than to 4 B, 0 Y.

We may venture the hypothesis that the stages we have distinguished in the development of the idea of a distribution partly represent logical 'structures' (in Piaget's sense) that emerge at different ages and partly represent the fruits of experience. These 'structures' undergo transformations during development. At first they are shaped by purely subjective preferences of one kind or another. These preferences gradually weaken and become displaced by factors inherent in the objective situation.

We should like to thank the Education Authorities at Manchester and Bolton and the principals and staff at their schools for their kind help in making it possible for us to carry out these studies.

(*Manuscript received* 1 *July* 1954)

THE SCIENTIFIC STATUS OF MENTAL MEASUREMENT

By F. W. WARBURTON

Department of Education, University of Manchester

I. *Psychology as an exact science* (pp. 122–124). II. *Scales of measurement in mental testing* (pp. 125–129). *References* (p. 129).

I. Psychology as an exact science

The nature of science is a matter of definition. Despite Kant's rejection of scientific psychology on the ground that mental states cannot be measured, Comte's objection that experimental methods could not be applied in psychology, and the pessimism of William James, psychologists have been in general agreement for over a hundred years that their subject is a science. They also agree that it is not an exact science, although if the exact sciences are defined as those that permit of precise quantitative measurement, it is clear that in certain areas psychology has already laid claim to be exact. The subject is commonly regarded as inexact in the more fundamental sense that there is imperfect agreement between the results of different investigations into the same problem, and that wide variations exist in individual and in group performance under similar experimental conditions.

This paper is largely concerned with the fields of mental measurement and personality testing. Its first purpose is to consider in turn certain factors making for variability of performance in different investigations, and to discuss very briefly the advantages and disadvantages of attempting to bring these factors under control. Six factors are considered: (i) the use of subjective judgements, (ii) control of the experimental situation, (iii) differences between the tests or indicators assumed to represent certain psychological variables, (iv) sampling error, (v) multidimensionality of the data, and (vi) variation within the individual.

The argument leads to a defence of psychometric techniques.

(i) An indispensable condition of an objective experiment is that the readings made by different observers agree. In other fields scientists may dispute the value of certain experiments or the techniques employed, but they do not disagree (in ordinary circumstances) about the actual readings. In statistical terms, inter-judge correlations should be plus one.

An objective psychology defined in this way will exclude two main types of data, first, aesthetic, moral and literary judgements, and secondly, assessments about which the observers fail to agree because the data are too complex to be fully understood. For example Binet testers will sometimes differ in their opinion as to whether a child understands the meaning of a word or the nature of a geometrical shape, or, in a reading scale, the testers may disagree about whether a child has read certain words correctly. The difficulty in these cases is not that the judges carry their own preferences with them, but that it is impossible to obtain an unambiguous answer to the question. The interpretation of many items in projection tests also falls into this category. Over a wide area of psychology, subjectivity of judgement springs from both these causes, the observers applying different standards to equivocal material.

The amount of unreliability in subjective judgements is extremely high. In the cognitive field we find that correlations between examiners usually fall into the 0·5–0·85

range. In personality assessment, inter-judge reliabilities are at much the same level, and average about 0·70. This means that, in terms of variance, a very considerable proportion of the information, often more than half, is due to the experimenter and not to the experiment. Observers also vary considerably in their own judgement from occasion to occasion, which adds to the overall unreliability. Exactitude in the experimental findings cannot be expected if the ability and personality of the investigator enters into the results.

(ii) Another major factor leading to disagreement between the results of different researches in the applied field is the impossibility of controlling the environment in real life situations. Assessments of a boy's personality in school A cannot be directly compared with those made in school B, since conditions in the two schools are different, and each estimate is relative to a different population. This introduces an additional element of unreliability which would be present even if the measures of the boy's personality at school were completely objective, such as the number of attendances at meetings of school societies. This limitation extends over much of the vocational field; it applies to all biographical material, and to follow-up data in the school, university or occupation. In the study of the behaviour of individuals in small groups also, there is no rigid control of the environment, since the members of the group differ from occasion to occasion. The same person may be loquacious in one group and silent in another. Greater agreement between the results of different investigations should come with increased standardization of the environment. It is well known, for example, that follow-up studies in the vocational field show greater agreement if the criterion is success on a training course, than if follow-up is made into success in a job. There should be still greater agreement if the training course is the same for all candidates, and so on. The only branches of psychology which seem likely to achieve exactness comparable with that found in the older sciences are accordingly those confined to the laboratory and a controlled environment, i.e. physiological, experimental and animal psychology, mental measurement, objective tests of personality and laboratory studies in social psychology. Together they constitute a considerable body of knowledge. As long ago as the eighteen-nineties Dilthey (see, for example, Hodges, 1952) suggested the division of psychology into two parts, an explanatory psychology that would imitate the natural sciences and attempt to discover the elements of conscious experience and the laws of their combination, and a descriptive psychology, which is an artistic process in which we are constantly aware of the relation of the parts to the totality. Many psychologists would regard any such distinction as artificial and an impoverishment of the subject. They take the view that psychology should not be divided according to the methods of analysis employed, but according to content, its data falling naturally into a few main branches, despite overlapping at the borderline.

(iii) A third source of inexactitude in psychological research is the use of different, and sometimes very small, samples of subjects to investigate the same problems. Experiments designed to reveal the influence of major factors such as sex and age may, in fact, yield conclusions which are true only in certain geographical areas or in certain social strata. This is another problem which is insoluble in the educational and vocational fields, although a greater attention to sampling differences both in the design of experiments and in the evaluation of the significance of the results, would help to bring more uniformity into research findings.

(iv) A fourth source of inaccuracy lies in the name or 'indication' given to variables. The Matrices and the Dominoes tests, for example, are both assumed to indicate non-

verbal intelligence, although the correlation between the two tests is considerably less than their reliabilities. In the personality field, different tests of the same trait often show very poor agreement, correlating at the 0·4 level. In some researches there is less relation between the same traits than between nominally different traits. From this point of view, factor analysis can be used, not to identify groups of similar tests, but to indicate the best single 'marker' i.e. the most typical test in a given field. These marker tests could be used in important investigations, non-verbal intelligence being replaced, let us say, by Matrices ability. Something of this sort happens by common usage when it is agreed that a certain measure will best serve our purpose, such as the Binet test in child guidance, the Wechsler-Bellevue in adult work, the Matrices in non-verbal intelligence, and the T.A.T. and Rorschach projection tests. Increased agreement between empirical results is likely to follow from agreement between the tests employed. These further conditions imply an even wider separation between a more exact psychology and everyday behaviour. The applied psychologist is building a mechanical man, whose characteristics are Matrix plus Stenquist plus Strong plus *X*.

(v) Another difficulty arises from the complexity of the subject-matter, the multi-dimensionality of the data studied by the psychologist is likely to result in further disagreement between the empirical findings. Guttman's scale analysis provides a valuable technique for isolating single dimensions, but once again the difficulty lies in extracting such dimensions from complex variables.

(vi) Variation in individual and group performance. This paper is concerned with differences in the results of different investigations, and not with test-retest reliability. It is well known, however, that variation in performance occurs even during a testing session. This factor will add to the disparity between the experimental results, just as snakes are more difficult to measure than carpets.

It is not suggested that exactness in the sense of precise agreement between the results of different experiments will come with the type of variable at present used in mental testing, even when identical, objective, unidimensional tests are given to representative samples in controlled situations. It is true that variation due to differences between judges and between tests may be removed by the use of identical objective tests. On the other hand, the control of the environment, of the experimental sample, and of the number of dimensions cannot be complete. Variation which remains after these rigid experimental controls have been applied falls into a different category. We do not seek to reduce this variation, but to gain an exact knowledge of its limits. It is for this reason that the theory of probability plays a large part in statistical psychology. If everyday behaviour is ever resolved in terms of static unfluctuating measures, it seems unlikely that they will be closely related to real life variables such as intelligence and dominance, although they might be physiological terms or resemble some of the constructs used in learning theory. The price to be paid for increased stability in experimental results is withdrawal from life, and opinion will differ as to whether the game is worth the candle. In any case, it is not to be expected that the postulates of measurement appropriate to the older sciences are necessarily relevant to psychology, in which the subject-matter is more complex. The primary aim of the psychometrist is not to reproduce the type of scale used in the physical sciences. Human beings do not behave like physical objects and it would be surprising if the same methods of measurement were efficient in both cases.

II. Scales of measurement in mental testing

The remainder of this paper consists in a reply to certain conventional criticism of the measures used in applied psychology, the critics taking the view that scores derived from mental tests are invalid since some of the fundamental principles of measurement are broken by psychometric techniques. These principles include the law of transitivity, which states that if A is greater than B and B is greater than C, then A is greater than C. In terms of the difficulty of test items this is not true for all individuals. A person can usually answer correctly some questions which are more difficult than those in which he fails. It is as if the same person were less than 5 ft. 8 in. tall but more than 5 ft. 9 in. The root difficulty arises from the complexity of the subject-matter. We are forced to measure multidimensional data in terms of unidimensional scales. Let us first of all take an example of a single dimension, which is measured in terms of a single scale (Table 1). Four persons are given a test comprising four questions: (i) Are you more than 5 ft. 4 in. tall? (ii) Are you more than 5 ft. 5 in. tall? (iii) Are you more than 5 ft. 6 in. tall? (iv) Are you more than 5 ft. 7 in. tall? These questions are plotted in order of difficulty and the persons in order of merit as follows:

Table 1

	Question				
Person	64 in.	65 in.	66 in.	67 in.	Total score
A	√	√	√	√	4
B	√	√	√	×	3
C	√	√	×	×	2
D	√	×	×	×	1
Total persons (difficulty)	4	3	2	1	

A answers 'Yes' to the most difficult question, i.e. he is more than 67 in. tall. Consequently, he also answers 'Yes' to the easier questions, since he is also more than 66, 65 and 64 in. Similarly, although *B* is not more than 67 in. tall, he is more than 66 in. and therefore also more than 65 and 64 in. and so on. The ticks and crosses in the table form solid blocks.

On the other hand, in Table 2, in which the answer plot resembles those obtained for mental tests, person *B* can answer question *S* correctly, but fails in question *R* which is less difficult. The intransitivity of the answer pattern follows necessarily from the multidimensionality of the data. To illustrate this, let us suppose that we are trying to measure a multidimensional variable, say a person's shape, which depends on height and breadth, in terms of a single dimension, the ability to pass through a series of doors. The test comprises a set of four doors, some narrow, some squat, some squarish, and so on, labelled *P*, *Q*, *R*, *S*, respectively. Let us also suppose that, like the mental tester, we do not know the exact nature of what we are measuring, although we are able to devise questions that appear relevant. We cannot distinguish between the height or breadth of either persons or doors, we simply know that different persons have different shapes and can pass through certain doors. If the questions are arranged in order of difficulty and the persons in order of merit we have the type of answer pattern shown in Table 2. The most difficult question is door *S*, since only one person can pass it. The easiest question is door *P*, which is passed by all four subjects.

Person *B* is tall and thin, so thin, in fact, that he is the only person who can squeeze through the narrow and difficult door *S*. On the other hand, his height prevents him from getting through door *R*, which is easier than door *S*. Person D is so fat that he is the only one who is unable to squeeze through the easy door *Q*, although he can get through the more difficult door *R* because he is short, and so on. The difficulty is that we are trying to reduce a multidimensional to a unidimensional scale, to express two variables in terms of one to measure height and width in terms of doors only. This type of answer plot in which the ticks and crosses are arranged higgledy-piggledy, is to be expected in mental tests.

Table 2

	Question (door)				
Person	*P*	*Q*	*R*	*S*	Total score
A	√	√	√	×	3
B	√	√	×	√	3
C	√	√	×	×	2
D	√	×	√	×	2
Total persons (difficulty)	4	3	2	1	

Another explanation may be advanced of this type of answer plot. It may be that the shapes of the subjects change while the experiment is in progress. They might, for example not be human beings at all, but amoeba. Consequently, a question or door which is difficult to a certain individual one moment may be easy the next. There is no way of telling from the answer pattern alone whether this unevenness is mainly due to the multidimensionality of the data or to the short-term variability of the individual.

The point of view taken in this paper is that the intransitivity of test measures, which comes largely from the use of multidimensional data, is unavoidable, regrettable, and of little consequence. In fact, intransitivity might well be handled satisfactorily by a branch of mathematics, such as the theory of groups, in which the formal requirements are less rigorous.

Two further criticisms of mental testing based on the theory of measurement are (i) that if measurement is used in psychology it should be in terms of linear units, which are constant in size, whereas mental tests are scored in terms of items of different difficulty, and (ii) that there is no true zero in mental measurement so that it is impossible to use ratios: the statement, for example, that one person is twice as intelligent as another has no meaning.

Various types of scale are possible, of which the most important to psychologists are: (i) ordinal scales, such as the scale of hardness for minerals; (ii) interval scales, which demand equality of intervals but do not require an absolute zero (Fahrenheit temperature scales are examples of this type of scale, in which ratios cannot be compared, e.g. a temperature of 212° F. is not twice as hot as a temperature of 106° F.); (iii) ratio scales, which demand equality of intervals and an absolute zero. Ratios may be compared, e.g. a person weighing 16 stone is twice as heavy as a person weighing 8 stone.

In psychology, interval and ratio scales play an important part in experimental and animal psychology. Reaction times, sensory thresholds, maze learning times and measures of fatigue, are chiefly of interest as absolute rather than ordinal measures. In vocational psychology, too, some occupational assessments such as size of vocabulary and shorthand speed are usually regarded as absolute rather than relative. For example, the number of

words per minute that a shorthand writer can take down is more important than her position in the commercial school.

The important point is that in the fields of mental measurement, personality testing and experimental social psychology, scales are rare in which some form of rank order is not clearly implied. Some measures of ability, such as I.Q.'s may become sufficiently well known to appear to be absolute, but an I.Q. of 160, like a size 17 shoe or a golfing score of 70, is relative to a population. We know from experience that it is uncommon. We rank it.

The applied psychologist is concerned with the distance between people in society, not between objects in space, he is more interested in rank order than in absolutes, and he therefore changes his raw scores into quotients, percentile ranks, mental ages, and scales based on the standard deviation, which indicate the relation of each measurement to other measurements in the same universe. The zero point is in the middle of the scale; it is the average score, and the scale extends on either side of it. To have 'zero' intelligence, in the ordinary sense of the term 'zero' is to be dead, and beyond the reach of the mental test. It is possible to build a scale of ability in terms of difficulty, which, although not unidimensional, has many of the same characteristics as the ordinary measuring rod, with a zero point and units which are additive and equal (Peel, 1949). Such scales are rarely constructed because they are not particularly useful. It would not be of outstanding importance (unless the measures were stable) if a cerebrometer were invented which measured absolute intelligence in equal and additive units of cerebral plasticity. The next step would still be to change the readings into some form of ordinal units. Critics of psychological measurement who insist on the use of interval and ratio scales might as reasonably criticize a geologist because he does not know the percentile score of a pebble that weighs one and a half ounces. The geologist would reply that he is not interested in the relationship between individual pebbles in a society of stones. This is particularly true in the field of personality, in which no universally accepted linear scale has been established, even by common usage. In the course of an account of the limitations of mental testing Chambers (1943) remarks that persons have been asked to give marks to their emotional moods on a scale ranging from zero to ten, and so to arrange the marks that they form a normal distribution. He asks whether this is so very far removed from the absurdity of treating love as a commodity which may be weighed and made into parcels and tied up with pink ribbon. It is an interesting commentary on the appropriate scale to use in this situation that when a young woman asks her young man how much he loves her, and he answers three and a half, or two million and two, his reply is accepted as persiflage, but when the discussion turns to rank order, things become grimmer. On hearing that he is not placed at the hundredth percentile, the young man may regard this scale as real enough, despite the fact that it is not based on linear and additive units, to take his life. The primary consideration is, in fact, not the type of scale used. The high jump and the mile race are measured in the ratio scales of space and time used by the physicist, but the underlying aptitudes are complex and the problem of human variability remains.

In short, in our present stage of knowledge, the variables used by the applied psychologist are multidimensional. It is difficult to think of any human characteristic, from religious conversion to the knee jerk, which is not multidimensional, and this limitation has to be accepted. We also have to accept the fact that human behaviour, at the level studied by the psychologist, fluctuates from occasion to occasion. It often happens, for example, that a person is, according to examination results, a good mathematician one day and a

poor one the next; an introvert at breakfast and an extrovert at dinner; his performance in the high jump may fluctuate within a single minute. This variability extends over the whole range of behaviour, and is unavoidable. Applied psychologists have to define their variables in terms of ratings and test scores, and to use these scores as their starting-point without making any assumptions about their underlying nature or the number of dimensions of the traits concerned.

Another line of criticism of the methods of measurement used in mental testing is derived from the holistic point of view, that to take over the mathematical methods of the physical sciences and to use them for the purposes of psychological description and analysis commits us to a fundamental flaw. We cannot add units of ability as we add twelve inches to make a foot; nor can we put samples of behaviour into a balance and subtract and multiply them as we subtract and multiply ounces and pounds. The psychologist, according to this view, is not justified in assessing traits and persons in the same way as the physicist measures lengths and weights: the variables with which he is concerned do not differ simply in numerical magnitude, they are qualities, not quantities. He assumes that mental tests are based on factors which can be combined or resolved like forces in accordance with the parallelogram law of dynamics, i.e. that they can be presented by vector addition. Mental characteristics, so the argument runs, cannot be treated in this way: they produce new and unpredictable properties in combination, resembling chemical properties not mechanical forces. An individual does not, it is argued, respond to a certain situation merely by using a particular quality in isolation, he reacts as a unified personality.

A similar criticism is that test scores are a kind of blurred average. Different people may give identical answers to a test, but use different methods of arriving at them. One person may visualize problems diagrammatically, another may rely on considerations of abstract logical relationships, and another fall back on the recollection of similar problems solved in the past. These individuals can be regarded as equal only in so far as they have gained equal marks on a particular test. They cannot be said to possess equal quantities of some such capacity as 'general intelligence', or 'manual dexterity'. The treatment of subjects with equal marks as of equal intelligence cannot, according to this view, be seriously maintained. With regard to this argument, Thurstone (1947, p. 54) considers that if critics were consistent who point out that identical test scores do not necessarily measure the same mental processes, they should also object to the statement that two men have identical incomes, for one of them earns while the other one steals. Burt (1940, p. 137) has pointed out that when the householder receives a bill from his grocer he does not argue that the pleasures of cheese and chocolate are commensurable neither with each other, nor yet with a magnum of champagne, and that consequently the addition exhibited in his bill contains a fundamental flaw.

The limitations of the method of analysis are common to science. In order to understand any situation we analyse it into elements and study the separate elements in isolation. This procedure is essential in any investigation of complex material. When the isolated elements are recombined, something is missing in the reconstructed situation, which was present in the original situation. Physical bodies are incompletely analysed into electrons and protons, and water is described, rather absurdly, as the combination of two gases. This difficulty is particularly marked if the subject-matter is as complex as human behaviour.

The analytic method has proved extremely useful to the applied psychologist, who, whilst stressing the limitations of analysis and the importance of studying the whole personality, also uses analytic concepts such as I.Q., aggression, reading age, distractibility and social status. It is difficult to see how he could proceed in any other fashion. The methods used at present are justified by the results: they have pragmatic sanction. Work in psychometrics, in vocational and educational selection, in social psychology and in the field of personality cannot stop because the methods of measurement are not the same as those used in the older sciences.

References

Burt, C. (1940). *The Factors of the Mind,* University of London Press.

Chambers, E. G. (1943). Statistics in psychology and the limitations of the test method. *Brit. J. Psychol.* xxxiii, 189–99.

Hodges, H. A. (1952). *The Philosophy of Wilhelm Dilthey.* London: Routledge and Kegan Paul.

Peel, E. A. (1949). Item difficulty as the measuring device in objective mental tests. *Brit. J. Stat. Psychol.* ii, 69–75.

Thurstone, L. L. (1947). Multiple Factor Analysis. University of Chicago Press.

BODY-CATHEXIS AND PERSONALITY

By SIDNEY M. JOURARD and PAUL F. SECORD

Department of Psychology, Emory University, Georgia

I. *Introduction* (pp. 130–131). II. *Measurement of body-cathexis* (p. 131). III. *Correlates of body-cathexis* (pp. 131–137): (i) *Body-cathexis, anxiety and security concerning the self* (pp. 131–132); (ii) *An interpersonal correlate of body-cathexis* (pp. 132–133); (iii) *Body size and body-cathexis* (pp. 133–134); (iv) *Body-cathexis and the ideal female figure* (pp. 134–135); (v) *Demonstration that the concept of ideal figure is shared* (pp. 135–136); (vi) *Sex differences in differentiation of body-image and self-concept* (pp. 136–137). IV. *Discussion* (p. 137). V. *Summary and conclusion* (p. 137). *References* (p. 138).

I. Introduction

The body has been studied by psychologists from a number of different viewpoints, but a curious omission has been the study of *feelings* about the body.

There can be little doubt that feelings about the body have marked behavioural consequences, as both casual and clinical observation attest. Thus, in our society one can readily note the amount of care and attention that persons devote to the grooming and modification of their body structure and appearance. From a clinical viewpoint it is a common observation that emotionally disturbed persons express concern about the appearance and functions of their body and its parts. Yet, reports of systematic empirical research into feelings about the body are rare.

An early interest in feelings about the body was expressed by the psychotherapists. Freud and Adler called attention to feelings of narcissism, anxiety, and inferiority, with the body and its parts as object. Schilder (1935) devoted a monograph to the role of the 'body image' in personality functioning and adjustment. But the method of inquiry followed by these workers was primitive; they made extensive generalizations about personality-in-general from the self-observations of patients in therapy.

Perhaps the earliest attempt to bring more sophisticated psychological method to bear on body feeling is provided by Schilder (1938). In his book *Psychotherapy*, he describes some questionnaires developed by him as an adjunct to group and individual therapy. These questionnaires served as a more efficient means of exploring the beliefs, feelings, and memories of patients undertaking therapy than the free interview. Schilder was interested in these data for reasons stemming from his theoretical concern with the body-image.

Machover (1948) has provided a 'projective' method of studying the body-image which utilizes drawings of the human figure. This method may be useful as a source of hypotheses concerning the single case, or as a method of confirming hypotheses deriving from other approaches, but it has not proven of much value for more systematic study.

Secord's word association procedure (1953) may be viewed as an improvement in methodology. He devised a set of 75 homonyms which could be perceived as words relating to the body, or as non-body-related words. The subjects were required to respond to the oral presentation of these words by writing down the first association that occurred to them. The words *colon*, *graft*, and *tablet* are illustrations. Three bodily responses might be: *colon:intestine*, *graft:skin*, *tablet:aspirin*; while three non-bodily responses might be: *colon:comma*, *graft:politics*, or *tablet:paper*. The total number of bodily responses for the

75 homonyms constituted the subject's score. Secord interpreted the number of body-related responses as indices of body-preoccupation, either narcissistic or anxious, and offered some tentative supporting evidence.

The present writers approach the problem of feelings about the body more directly. They raised the question: 'Can feelings about the body be quantified?' The problem seemed to lend itself nicely to methods already familiar in questionnaires and rating scales. Accordingly, they defined one variety of feeling, body-cathexis—the degree of satisfaction with the parts of the body—and formulated a questionnaire for its direct measurement. Parenthetically, it should be noted that the present use of the term differs from the psychoanalytic meaning which indicates the amount of instinctual energy invested in an object. In the present paper, the term refers to a continuum of feeling ranging from strong positive to strong negative.

II. The measurement of body-cathexis

The most recent Body-cathexis (BC) scale is a listing of forty parts or aspects of the body which are rated by subjects according to the following five-point scale:

(1) Have strong positive feelings.
(2) Have moderate positive feelings.
(3) Have no feeling one way or the other.
(4) Have moderate negative feelings.
(5) Have strong negative feelings.

Thus, if a person were very dissatisfied with his weight, he would encircle a (5).

The forty body parts are: hair, facial complexion, appetite, hands, distribution of hair (over body), nose, physical stamina, elimination, muscular strength, waist, energy level, back, ears, age, chin, body build, profile, height, keenness of senses, tolerance for pain, width of shoulders, arms, chest (or breasts), appearance of eyes, digestion, hips, resistance to illness, legs, appearance of teeth, sex drive, feet, sleep, voice, health, sex activities, knees, posture, face, weight, sex organs.

Although it is subject to some of the pitfalls and shortcomings already discovered in questionnaires or self-inventories this method of study has the advantage of relative simplicity, reliability and much previous work done with it.

In some of the studies to be reported, a self-cathexis (SC) scale similar in nature to the BC scale was employed. Self items ranged from 40 to 56 in the various forms of this scale. Examples of self-cathexis items are: morals, ability to express self, artistic talents, intelligence level, conscience, personality, self-discipline, vocabulary.

III. Correlates of body-cathexis

(i) *Body-cathexis, anxiety and security concerning the self*

The first study (Secord & Jourard, 1953) undertaken with this instrument (hereafter called the BC scale, for body-cathexis) was conducted with a total of 126 college male and female subjects. The BC scale was submitted to 70 males and 56 females along with a 56-item SC scale, and the homonym test of anxious body concern (Secord, 1953). Some

cases were eliminated for reasons of 'response-set'. All three scales were demonstrated to possess adequate reliability, even with the exclusion of these cases.

The following results were obtained, as shown in Table 1.

Table 1. *Intercorrelations among BC-SC scores, the homonym test, and the Maslow test*

Test	Body-cathexis	Self-cathexis	Anxiety-indicator
Honomym test			
Males ($N=45$)	−0·18	−0·15	−0·37*
Females ($N=43$)	−0·41**	−0·23	−0·40**
Maslow test ($N=46$)	−0·37*	−0·52**	−0·41**
Self-cathexis			
Males	0·58**		
Females	0·66**		

* 0·02 level of significance. ** 0·01 level of significance.

Significant correlations were obtained between BC and SC for both sexes, between the homonym test and total BC for females but not for males; and for both sexes between the homonym test and an eleven-item '*anxiety-indicator score*' derived from the BC scale.[1]

A subsequent study with a mixed group of 47 males and females yielded a significant correlation between the Maslow test of Psychological Security-Insecurity (Maslow, Hirsch, Stein & Honigmann, 1945), and BC, SC, and the 'anxiety-indicator score'. These data also appear in Table 1. The Maslow test is of the self-inventory variety, containing 50 items to be answered by *yes*, *no*, or *?*. Sample items are: 'Do you ordinarily like to be with people rather than alone?' 'Do you lack self-confidence?' These items were selected by Maslow from a large number on the basis of the extent to which they discriminated between normals and between persons characterized as insecure on the basis of clinical interviews.

From the results in Table 1 it is clear that body-cathexis is a variable which can be measured, and which is meaningfully related to other variables with an adjustment connotation, namely self-cathexis (the degree of satisfaction-dissatisfaction with aspects of the self), security with respect to the self (as measured by Maslow's scale), and Secord's projective-homonym test of anxious body-concern.

(ii) *An interpersonal correlate of body-cathexis*

The above study demonstrated that different degrees of body-cathexis 'make a difference'; i.e. that BC is indeed a 'significant' variable in personality research and theory. The next research to be described was concerned with the question: 'Will a person's attitudes and feelings about his body be related to attitudes believed to be held by "significant others"?'

Remy (1953), a student of the authors', conducted a study on this problem. Using 48 college males and 51 college females as subjects, he had them fill out three BC-SC questionnaires: first, according to their own feelings; secondly, as they believed their mothers felt about their (the S's) bodies and selves; and thirdly, as they believed their fathers felt. Finally, the subjects filled out a Maslow scale.

The main results of this study are summarized in Table 2. It may be seen that *perceived* mother and father attitudes about the S's bodies and selves correlate to a significant

[1] The 'anxiety-indicator' score was based on the sum of the cathexis-ratings assigned by each subject to the eleven items that received the greatest number of negative ratings by the group.

degree with self-evaluations, and that both parental cathexes, and BC-SC scores correlate with the Maslow test of security (Mas).

Table 2. *Intercorrelations among BC-SC scores, perceived parental cathexes, and security scores*

	Self-rating		Perceived mother-ratings		Perceived father-ratings	
Test	BC	SC	BC	SC	BC	SC
BC	—	0·68**	0·68**	—	0·56**	—
	—	0·84**	0·74**	—	0·68**	—
SC	—	—	—	0·77**	—	0·66**
	—	—	—	0·70**	—	0·65**
Mas	−0·50**	−0·66**	−0·57**	−0·63**	−0·42**	−0·49**
	−0·37*	−0·63*	−0·32*	−0·37*	−0·33*	−0·37*

The first r in each cell is for $N = 51$ females; the second r is for $N = 48$ males.
* Significant at the 0·05 level of confidence.
** Significant at the 0·01 level of confidence.

The data obtained by Remy signify that within the 'psychological world' of the subjects, if they believe that their parents dislike assorted aspects of body and self, then they hold mainly similar body-self attitudes, and appear to be generally insecure, in so far as the latter is tested by Maslow's test. If one might assume that the subjects are able to estimate their parent's feelings with some accuracy, certain theoretical propositions outlined by Sullivan (1947), Murphy (1947) and Rogers (1951), with respect to the correlation of self-appraisals with the appraisals of significant others are supported by the data. Some knowledge concerning the accuracy of the estimates is needed, however, before any definite conclusions can be drawn.

(iii) *Body size and body-cathexis*

The next studies to be reported are concerned with the question: 'Why will a subject report that he likes or dislikes a given body part?' Two separate researches support the contention that the *measured size* of certain body aspects is an important determiner of the subject's feelings about those body parts.

The first of these studies (Jourard & Secord, 1954) was conducted with 62 college males. Subjects first filled out the BC-SC scale. Without prior warning, measurements were then taken of height, weight, width of shoulders, circumference of chest when relaxed and expanded, and circumference of biceps when relaxed and contracted. Correlations were computed between these measurements and the five relevant BC ratings: *height*, *weight*, *shoulder width*, *chest* and *muscular strength* (see Table 3).

Table 3. *Correlations between size and cathexis for selected body aspects*

($N = 62$ males.)

Body-cathexis	Actual size	r
Height	Height	−0·55**
Weight	Weight	−0·04
Shoulders	Width of shoulders	−0·36**
Chest	Chest (relaxed)	−0·33**
Chest	Chest (expanded)	−0·37**
Muscular strength	Biceps (relaxed)	−0·52**
Muscular strength	Biceps (contracted)	−0·54**

** $r = 0{\cdot}33$ at the 0·01 level for D.F. = 60.

All of the correlations, except that with weight, while low, are significant. They indicate, from their direction, that *large* size of the relevant body parts is associated with positive cathexis, while the reverse is true for small size. It was suggested in the report of this research that these data reflect the cultural stereotype or ideal with respect to masculinity; namely, that in our culture, *it is important and masculine to be big.*

The reverse trend was found with females (Jourard & Secord, 1955). Cathexis-ratings on a seven-point scale and direct measurements were taken of the following twelve body parts of 60 college females: *height, weight, bust, waist, hips, thighs, calves, ankles, feet, nose length, shoulder width* and *neck length.* Correlations between these pairs of measurements are shown in Table 4. All but four of these correlations between size and cathexis ratings were significant. The direction of these *r*'s suggested that *small size* was most desirable to the subjects, with the exception of *bust*, where large size was desired.

Table 4. *Correlations between cathexis-ratings and measured size of body parts*

($N = 60$ females)

Height	0·19	Calves	0·39**
Weight	0·37**	Ankles	0·46**
Bust	−0·25*	Feet	0·52**
Waist	0·32*	Nose length	0·22
Hips	0·43**	Shoulders	−0·13
Thighs	0·42**	Neck length	−0·17

* Significant at the 0·05 level. ** Significant at the 0·01 level.

The findings with males and females both implied, without demonstrating, the presence of cultural ideals with respect to masculinity and femininity. This ideal was not studied among the males, but was explored extensively among the group of sixty females described above.

(iv) *Body-cathexis and the ideal female figure*

The sixty females reported above were asked to give *estimates* of the size of the twelve body parts, and also to specify the dimensions which they believed would be most ideal for them. The hypothesis governing this aspect of the study may be stated as follows: cathexis for a body part is a function of conformity with or deviation from the subjective ideal; the more closely a girl's body approximates her concept of ideal dimensions, the more she will be satisfied with this body part.

Accordingly, deviation scores were computed for the five body parts most likely to be known by the *S*'s; viz.: *height, weight, bust, waist* and *hips.* Two deviation scores were obtained, one, by subtracting the actual measurements from the ideal dimensions, ignoring sign, and another, by subtracting the estimated measurements from the ideal. These may be regarded as measures of objective, or 'real' deviation from the ideal, and perceived deviation from ideal, respectively. As shown in Table 5, both sets of scores correlate quite highly with cathexis-ratings for the relevant body parts.

Table 5. *Correlations between cathexis-ratings and deviations from ideal size for selected body parts*

(The two discrepancy scores are obtained by subtracting the measured size and estimated size, respectively, from self-ratings of ideal size ($N = 60$ females).)

Discrepancy score	Height	Weight	Bust	Waist	Hips
Ideal-measured size	0·41**	0·50**	0·11	0·42**	0·32*
Ideal-estimated size	0·55**	0·61**	0·21	0·50**	0·49**

* Significant at the 0·05 level. ** Significant at the 0·01 level.

Another observation may be made with respect to the nature of the ideal—an observation consistent with the correlations shown in Table 4. In three body aspects (*weight*, *waist* and *hips*) the self-rated ideal is significantly smaller than the measured size; the ideal for *bust*, on the other hand, is significantly larger than the average bust measurement for the group. The data on which these observations are based are shown in Table 6.

Table 6. *Comparisons of mean difference between measured (and estimated) size and ideal size*

($N = 60$ females)

Body part	Measured size	Ideal size	Mean diff.	$\sigma_{\text{mean diff.}}$	C.R.	*P*
Height	65·69 in.	65·53 in.	0·16	0·20	0·80	>·05
	Est. *65·52 in.*	—	*0·01*	—	—	*>·05*
Weight	125·42 lb.	122·48 lb.	2·94	1·49	1·98	·05
	Est. *126·97 lb.*	—	*4·49*	*1·35*	*3·32*	*<·01*
Bust	34·14 in.	34·83 in.	0·69	0·20	3·45	<·001
	Est. *34·28 in.*	—	*0·55*	*0·19*	*2·89*	<·01
Waist	25·45 in.	24·27 in.	1·18	0·18	6·56	<·001
	Est. *25·57 in.*	—	*1·30*	*0·18*	*7·22*	*<·001*
Hips	37·43 in.	35·06 in.	2·37	0·29	8·17	<·001
	Est. *36·84 in.*	—	*1·78*	*0·23*	*7·74*	*<·001*

Est. = estimated size.

The average deviations from ideal measurements may be described in terms of a hypothetical average girl (in the sample). She is 2·94 lb. overweight, has a bust measurement which is 0·69 in. too small, a waist that is 1·18 in. too big, and hips that are oversize by 2·37 in.

These differences between measured and ideal mean sizes are small. Yet, when we are dealing with body measurements, even these slight differences are highly significant; phenomenologically, the small differences truly make a difference.

(v) *Demonstration that the concept of ideal figure is shared*

Each girl in the study being reported was asked to specify the measurements which would be ideal *for her*. If these ideal measurements were unshared, we might expect them to vary within the sample as widely as direct measurements of body parts; i.e. there would be as many different ideals as there were objective body proportions. If, on the other hand, these ideal concepts are shared as a sort of cultural stereotype, we would expect the variation among ideal measurements to be smaller than that for direct measures. Accordingly, standard deviations were computed for each ideal measure, and these were compared with standard deviations for measurements of the five body parts. As shown in Table 7, the standard deviations for ideal measurements are all significantly smaller than

Table 7. *Comparison of* S.D.*'s for measured size with* S.D.*'s for ideal size*

($N = 60$ females.)

Body parts	Standard deviations (S.D.'s)						
	Measured size (x)	Ideal size (y)	r_{xy}	$\sigma_x - \sigma_y$	$\sigma_{\text{diff.}}$	C.R.	*P*
Height	2·12	1·63	0·68	0·49	0·18	2·72	<0·01
Weight	16·79	9·17	0·76	7·62	1·25	6·10	<0·001
Bust	2·06	1·31	0·61	0·75	0·18	4·17	<0·001
Waist	1·71	1·15	0·62	0·56	0·15	3·73	<0·001
Hips	2·20	1·56	0·33	0·64	0·23	2·78	<0·01

those for measurements of the actual size of the body parts. Apparently some factor, probably the culturally defined and shared concept of the ideal, restricts the range of variation of acceptable and desirable body size to a much smaller margin than, say, appetite and given bodily endowment allow. This finding surely has implications for personal adjustment, for it signifies that the cultural ideal of female bodily beauty is one that is difficult, if not impossible to attain; thus, many females are doomed to insecurity by the very restrictive nature of the ideal.

(vi) *Sex differences in differentiation of body-image and self-concept*

An interesting question concerns the extent to which an individual assigns similar ratings to all body parts, or, on the other hand, varies widely in rating different body parts. Thus, at the extremes, he may have a 'global' or a 'differentiated' concept of body and self.

Remy (1953) conducted some exploratory analyses of differentiation of the body-image and the self-concept in which he computed individual variances for BC and SC ratings. He found that females give a higher mean variance than males in responding to the BC scale, but the sexes did not differ in mean variance for SC ratings. The males also had a significantly lower mean variance for BC ratings than they did for SC ratings, while the women did not differ in differentiation of body and self.

The relevant analyses and comparisons are shown in Table 8.

Table 8. *Difference between mean variances for self-rated body-cathexis and self-cathexis in male and female subjects*

($N = 48$ males and 51 females.)

	Between-sex comparisons			
	Mean variance			
Scales compared	Males	Females	Difference	t
Body-cathexis	0·76	0·96	−0·20	2·13*
Self-cathexis	1·01	1·02	−0·01	—

	Within-sex comparisons		
	Mean variance	Difference	t
Males			
Body-cathexis	0·76	−0·25	3·37**
Self-cathexis	1·01		
Females			
Body-cathexis	0·96	−0·06	—
Self-cathexis	1·02		

* Significant at the 0·05 level. ** Significant at the 0·01 level.

These findings suggest that women in our culture make finer cathexis discriminations within their phenomenal body structure than males. Perhaps the greater variability shown by women in assigning cathexis-ratings to their body parts reflects a cultural limitation on the means of attaining security and acceptance from others[1]—a woman's body is somewhat more important to her attainment of acceptance and security than a man's body is to a man. Men in our culture may have more avenues of attaining security through traits of self. The higher correlation among women than among men between

[1] Murphy points out (1947, p. 345) that differentiated perceptions emerge from more global ones under the impetus of need-tensions.

BC and security scores which Remy found (see Table 2) is perhaps additional evidence in support of the stated propositions concerning the greater personal and social importance of the female body.

IV. Discussion

From the data reported herein, some general comments about the role of the phenomenal body in personality may be offered.

We can suggest that a person's security—his felt self-acceptability as well as his acceptability to significant others—is related in some way to his feelings about his body. Perhaps the explanation of this relationship is that a person in our culture perceives his body appearance as a tool for impressing others, whose approval he has come to need for his self esteem. If his body does not correspond with his standards (internalized from others), he will feel insecure.

This interpretation is supported by the findings with respect to body size and body-cathexis. Men and women alike, it appears, acquire a concept of ideal body proportions, and vary their cathexis ratings with the degree to which they approximate these ideal specifications.

With the data already at hand, these considerations seem to apply most strikingly to females, although there is evidence to indicate that they apply to males as well.

There remain many unsolved problems relating to body-cathexis. Two important ones which have not been studied concern the origins of the body attitudes, and the change of cathexis-ratings of body parts with time.

The method which has been described should, with suitable modifications, be useful in the solution of some of these problems.

V. Summary and conclusion

Body cathexis is defined as the degree of satisfaction reported by an individual for the parts of his body. A method for measuring this variable has been described, and the results of a number of studies reported. The findings to date may be summarized as follows:

1. Body-cathexis measures are correlated with measures of self-cathexis, anxious body-preoccupation, and security concerning the self.

2. Body- and self-cathexis measures are correlated with measures of perceived parental attitudes to the body and the self; parental attitudes to the body and the self are also related to security concerning the self.

3. The measured size of the body is an important determiner of cathexis for the body. With males, large size appears to be correlated with positive cathexis, whereas with females, small size seems to be desirable.

4. Among females, self-rated ideal size is smaller for most body parts than the measured size; one exception is for *bust*, where the ideal size is larger than the measured size.

5. The concept of ideal size, among females, appears to be a shared, or stereotyped concept of the ideal, and it shows much less variability within the group than measured size.

6. Variability in assigning cathexis-ratings to the parts of the body is greater among females than among males. The sexes do not differ in variability in rating the self. Among males, greater variability is shown in rating the self than the body; the women do not manifest this difference.

The implications of these findings for personality adjustment are briefly discussed.

References

Jourard, S. M. & Secord, P. F. (1954). Body size and body-cathexis. *J. Cons. Psychol.* xviii, 184.

Jourard, S. M. & Secord, P. F. (1955). Body-cathexis and the ideal female figure. *J. Abnorm. (Soc.) Psychol.* (in the Press).

Machover, Karen (1948). *Personality Projection in the Drawing of the Human Figure.* Springfield: C. C. Thomas.

Maslow, A. H., Hirsh, E., Stein, M. & Honigmann, I. (1945). A clinically derived test for measuring psychological security-insecurity. *J. Gen. Psychol.* xxxiii, 21–41.

Murphy, G. (1947). *Personality; a Biosocial Approach to Origins and Structure.* New York: Harper.

Remy, R. (1953). Relationships among Body- and Self-Cathexes, Perceived Parental Cathexes, and Security. Unpublished master's thesis, Emory University.

Rogers, C. (1951). *Client-Centred Therapy.* Boston: Houghton-Mifflin.

Schilder, P. F. (1935). *The Image and Appearance of the Human body.* London: Kegan Paul.

Schilder, P. F. (1938). *Psychotherapy.* New York: Norton.

Secord, P. F. (1953). Objectification of word association procedures by the use of homonyms: a measure of body-cathexis. *J. Personality,* xxi, 479–95.

Secord, P. F. & Jourard, S. M. (1953). The appraisal of body-cathexis: Body-cathexis and the self. *J. Cons. Psychol.* xvii, 343–7.

Sullivan, H. S. (1947). *Conceptions of Modern Psychiatry.* Washington: Wm. Alanson White Fdn.

(*Manuscript received* 26 *February* 1954)

THE DVORINE PSEUDO-ISOCHROMATIC PLATES[1]

By AGNES CRAWFORD

Department of Psychology, University of Liverpool

In 1944, Israel Dvorine published two volumes of colour plates; vol. I he called *Colour Perception Testing Charts for Ages 7 and Over*—a diagnostic set of pseudo-isochromatic charts, and vol. II he called *Colour Perception Training Charts* (Dvorine, 1944*a*). These were published by the author himself and described as available from the author, at a cost of $25.00 (Andrews, 1948). In the same year and in 1945 and 1946, Dvorine published a series of papers with titles as follows: 'A new diagnostic method of testing and training colour perception' (Dvorine, 1944*b*), 'Reconditioning the colour blind; a case report' (Dvorine, 1944*c*), 'The use of volume two of the Dvorine colour perception testing and training charts for grading colour perception difficulties' (Dvorine, 1945*a*), and 'Improvement in colour vision in twenty cases' (Dvorine, 1946).

Papers with such titles could not long pass unchallenged, nor could his claims for the tests. Elsie Murray reviewed the tests in 1945 (Murray, 1945*a*), and was unfavourably impressed both by the plates and by the claim that the charts could be and ought to be used for training colour perception. Her paper led to a somewhat acrimonious though largely irrelevant reply by Dvorine (Dvorine, 1945*b*) and to further criticism by Murray (Murray, 1945*b*). A few other cursory reviews appeared, but the only psychologist interested in colour vision testing at that time who seems to have used the new test at all was Louise L. Sloan. In a comparative study of some tests of colour vision (Sloan, 1945*a*), she used eleven plates of the Dvorine test and quotes the frequencies of scores made by a group of 100 subjects with normal colour vision and by a group of forty-seven subjects with colour-defective vision. She did not give much information about the test, but seems to have selected the eleven plates which her results showed to be most discriminating between previously classified groups of subjects. The table she gives shows no overlap between the scores obtained on these eleven plates by the colour-normal and the colour-defective subjects.

Dvorine continued to write about the tests (Dvorine, 1948). It was not, however, one of the tests selected by Chapanis (1948) for his comparative study of five tests of colour vision, so that, presumably, he was not impressed by the test, nor by its author's claims.

It is clear that the first edition of the Dvorine test was not orthodox, and Louise L. Sloan seems to have been the only person to take it seriously and to use parts of it in the orthodox way. Now, in 1953, we have what is called a second edition of the test—really, however, a new version of the test.

The first page of this new version of the test consists of a most handsome acknowledgement by Dvorine to Dr Louise L. Sloan, and an admission that, starting from her paper in 1945, he had taken the eleven plates which she found to be highly diagnostic, and had used these as a 'nucleus for the second edition'. He has added some new plates and tested them on groups of colour-normal and colour-defective subjects, retaining only those plates

[1] Second edition, 1953, published by and available from the author at Eataw Place, Baltimore, Md. No price given.

which as he says, give a 'high percentage of normal subjects passing and colour deficients failing this test'. This, of course, is a criterion suggested by Sloan (1945*b*). Dvorine does not say on what criterion, other than the test itself, the subjects were classified as normal or colour deficient, nor how many subjects he used.

In the 'Instructions', for giving the test, Dvorine describes it as consisting of two parts: Section One contains, he says, 'One demonstration plate (the number 48 in red on a blue background) and 14 plates made up of eight different colour combinations arranged in pairs of similar colours. Section Two contains one demonstration plate (a blue trail on a red background) and 7 plates featuring trails instead of digits. . . the second section may be used to test pre-school age children and illiterates, or as a corroborative test' for the first section. He goes on to give details about the light in which the test should be given, the distance from the subjects' eyes, and to say that no more than five seconds should be allowed for each plate. These are points on which he was criticized by Murray (1945*a*).

On a third page, Dvorine gives what he calls 'Analysis' of the results of testing. This is now quite orthodox. Three or more failures on the fourteen digit-plates indicate red-green colour deficiency. Plates 6 and 7 are diagnostic for 'protanoid' and 'deuteranoid' types of colour deficiency. Two errors or one error can be allowed because of what he calls 'figure-ground confusions by otherwise normally seeing individuals'. Though he has, he says, studied such errors carefully, he cannot account for them. The same sort of error, in fact, appears on the Ishihara test, as any item analysis will show, and it should not perhaps be dismissed as easily as Dvorine suggests.

There is, however, a third section of the test, which Dvorine does not mention and which presumably remains as a legacy from his unregenerate days. It appears in the sample Record Form as a test of Nomenclature, and consists of a movable disk inside a double page; on the disk there are circles of colour which can be shown in succession through an aperture. The colours are what he calls on the Record Form, 'light and dark tints of Red, Brown, Purple, Yellow, Blue, Green, Gray and Orange'. The colours on the digit plates seem much purer and clearer than those on the colour-naming disks.

The test, obviously, has a more varied history than tests of colour vision usually have. A copy of the new edition was given to me for review, and the only way to reach a justifiable opinion about it was, clearly, to try it out, along with a well-established test, and a criterion test. Accordingly, the Dvorine test, the Ishihara test (Lewis's printing, 9th edition), and Rayleigh equations on an anomaloscope (Crawford, 1951*a*, *b*) were used, with as many subjects as could be mustered in the time available. Seventy-six subjects were obtained, mainly from the first-year course in psychology here, though this number includes a number of known colour-blind individuals who volunteered or were persuaded to come for test. In all, sixty-five subjects (forty-one women and twenty-four men) had normal colour vision on both the Plate tests, and eleven (ten men and one woman) were colour defective on both of these tests. Some of the subjects in the 'normal' group whose matches on the Rayleigh equation were at the extremes of the distribution could perhaps be regarded as very slightly deviant, but they show no overlap on any of the tests with those whom I have grouped as colour blind. Only the woman subject and one of the colour-blind men did not previously know or suspect that their colour vision was at least different from that of other people.

I propose to deal mainly with the results of the Dvorine test here, since the main purpose of the investigation was to find out how good or how bad this new version of the test was.

For most of the subjects, the first part of the test only was used. About one-third of the subjects did the second part of the test—i.e. tracing out the wavy lines or trails—sufficient to show that the results for that section will be very similar to those obtained from the first part. The colour-naming section was omitted altogether, after a few trials. The colours are unsatisfactory, and the disks seem to provide a test of the subjects' knowledge of present-day fashionable colour names, rather than of simple colour naming.

The scores for the sixty-five colour-normal subjects and the eleven colour-defective subjects on both of the pseudo-isochromatic tests are as follows:

Dvorine test		Ishihara test		
No. of plates read correctly	Frequency	No. of plates read correctly	Frequency	
14	51	20	18	
13	10	19	22	
12	4	18	10	
11	1 (colour anomalous man)	17	8	
5	1 (colour anomalous woman)	16	4	borderline scores
4	1	15	3	
1	2	14	1	(colour anomalous woman)
0	6	8	1	
		3	1	
		0	8	
	Total 76		Total 76	

It will be seen that the Dvorine test compares favourably with the Ishihara test. For this, the score is based on the twenty figure-plates which can be read by those with normal colour vision. There is one subject—anomalous on the anomaloscope matchings—who has a borderline score on the Dvorine test, compared with seven or eight on the Ishihara test, only one of whom was clearly defective on the anomaloscope matchings. No one who is clearly defective on the Ishihara test has a better score on the Dvorine test, but some who are 'normal' on the Dvorine test (and on the anomaloscope matchings) have questionable scores on the Ishihara test.

An item analysis of the distribution of errors on the Dvorine plates shows a slight tendency for the relatively small number of errors—(30 errors by sixty-five normal colour-vision subjects)—not to be evenly distributed over the fourteen plates, but to accumulate on plates, 2, 3, 4 and 6, though only to the extent of 6 or 4 misreadings, instead of just over 2 that might be expected with this number of misreadings. This, however, is not such an uneven distribution as is found in any item analysis of the Ishihara test (Crawford, 1950; or Pickford, 1950). With some of the less markedly defective subjects who would be found in a larger sample, it might be necessary to make some allowance for these more frequently occurring errors, though the figures available are not sufficiently large to show whether this divergence from an expected even spread of errors over all the fourteen plates is statistically significant or not.

Correlation with the Ishihara test cannot be calculated, nor, with distributions of the kind shown here, could any correlation be calculated with the anomaloscope measurements. However, it is clear from the two frequency tables, that there are no gross misclassifications. The eleven subjects at the bottom of the tables are classified as colour blind by both the tests. It seems, from the present evidence that the Dvorine test discriminates

between the colour normal (by the criterion of the Rayleigh equation matchings on the anomaloscope) and the colour defective (by the same criterion). Colour normals do obtain scores of 12 or more out of 14; colour defectives score 11 or less. With a larger sample there is, of course, the possibility that some subjects reading eleven plates normally might also give normal matchings on the anomaloscope, but the proportion of the present sample giving this borderline performance is sufficiently small to justify the suggestion that such cases could be tested, for confirmation, by some other test. On the Ishihara, there are eight subjects who would need further testing to be sure whether their colour vision was normal or not.

Dvorine claims that the test will allow a diagnosis of 'protanoid' or 'deuteranoid' colour defect to be made. This, however, is not so clear-cut as he suggests. Out of the eleven colour defective subjects, only three were clearly classified; these were 'protanoids', and both the Ishihara and the anomaloscope matchings support this classification. The eight others could not be clearly classified, either because they failed to read both of the diagnostic figures on both plates 6 and 7, or because they read both figures on these plates, with a slight error in one only, or with some hesitation. The Ishihara classifies some of these as deuteranopes, but the anomaloscope matchings put them as deuteranomalous rather than as deuteranopes. One unclassified subject appeared as a definite protanope on the anomaloscope. In general, it might be said that the classification appears to be effective for well marked protanopes, but not for deuteranopes, and the test does not distinguish the anomalous trichromats from the dichromats. However, it may be that Dvorine is asking too much of a printed pigment test in expecting that it should be able to make this finer classification.

Apart from the failure to live up to this last claim, the new version of the Dvorine test seems to me to be an effective screening test, to separate the colour normal from the colour defective. The relatively small number of people who would give borderline scores (i.e. eleven plates read correctly) compares very favourably with the eight people who even on this small sample, would have borderline scores on the Ishihara test. While it is perhaps satisfactory from the point of view of test construction that eight out of eleven colour-defective subjects should have scores of 1 or 0 on the test, it is a little unsatisfactory from the subject's point of view. They are shown one demonstration plate with a clear figure on it, and asked to find other figures of the same sort, and then presented with a series of fourteen plates not one of which appears to them to have a figure on it. Apart from the feeling of resentment aroused by the notion that some trick is being played on them, it seems to me that this might, especially with younger subjects, arouse considerable anxiety about their 'eyesight' in general. On the method of scoring the Ishihara in terms of plates which normals can read, there are four plates which colour-blind people can read, which are not counted in the score, and there are eight or ten other plates on which the colour-blind subjects can see alternative figures. If the Dvorine test were being widely used, it might be advisable to have available some of these plates from the Ishihara test, to insert at judiciously chosen moments in the testing session, so that the subject is not too puzzled and worried by his apparent failure to do what the examiner is expecting him to do. This, however, is not a problem of colour vision, but a problem of testing technique.

REFERENCES

ANDREWS, T. C. (1948). Some psychological apparatus. A classified bibliography. *Psychol. Monogr.* LXII, 289.

CHAPANIS, A. (1948). A comparative study of five tests of colour vision. *J. Opt. Soc. Amer.* XXXVIII, 626–49.

CRAWFORD, A. (1950). An investigation of colour blindness. Ph.D. Thesis, Library of the University of London.

CRAWFORD, A. (1951*a*). A description of an anomaloscope. *Brit. J. Physiol. Optics*, VIII, 173–5.

CRAWFORD, A. (1951*b*). Rayleigh equations on an anomaloscope. *Brit. J. Physiol. Optics*, VIII, 225–34.

DVORINE, I. (1944*a*). *Colour Perception Testing Charts for Ages 7 and Over* (vol. I), and *Colour Perception Training Charts* (vol. II). Baltimore: The Waverly Press, $25.00.

DVORINE, I. (1944*b*). A new diagnostic method of testing and training colour perception. *Amer. J. Optom.* XXI, 225–35.

DVORINE, I. (1944*c*). Reconditioning the colour blind; a case report. *Amer. J. Optom.* XXI, 508–10.

DVORINE, I. (1945*a*). The use of volume two of the Dvorine colour perception testing and training charts for grading colour perception difficulties. *Amer. J. Optom.* XXII, 478–9.

DVORINE, I. (1945*b*). Author's comments upon the review of colour perception charts. *Amer. J. Psychol.* XLVIII, 397–9.

DVORINE, I. (1946). Improvement in colour vision in twenty cases. *Amer. J. Optom.* XXIII, 302–21.

DVORINE, I. (1948). Dvorine colour discrimination screening test. *Amer. J. Optom.* and *Arch. Amer. Acad. Optom.* XXV, 130–3.

MURRAY, E. (1945*a*). Review of Dvorine's colour perception charts, vols. I and II. *Amer. J. Psychol.* XLVIII, 276–9.

MURRAY, E. (1945*b*). A reply to Dvorine's comments. *Amer. J. Psychol.* XLVIII, 399–402.

PICKFORD, R. W. (1950). An item analysis of the Ishihara Test. *Brit. J. Psychol.* XLI, 52–62.

SLOAN, L. L. (1945*a*). An improved screening test for red-green colour deficiency composed of available pseudo-isochromatic plates. *J. Optical Soc. Amer.* XXXV, 761–6.

SLOAN, L. L. (1945*b*). Two criteria for the selection of colour vision test plates. *J. Opt. Soc. Amer.* XXXV, 723–30.

(*Manuscript received* 2 *August* 1954)

PUBLICATIONS RECENTLY RECEIVED

Method and Theory in Experimental Psychology. By C. E. Osgood. New York and London: Geoffrey Cumberlege, Oxford University Press. 1953. Pp. viii + 800. 80*s*.

The striking characteristics of this large book are these: it is written by a middle-of-the-road psychologist. It adheres to the classical quadri-partite division into sensory, perceptual, learning and symbolic processes. It seeks to cover the whole of this field. It has a bibliography of some twelve hundred titles.

The author's aim has been to provide 'a text that evaluates experimental literature in close relation to critical theoretical issues' giving 'a more detailed description of critical experiments' than is usual, and attempting 'to provide an interpretative framework that both brings out the significance of existing research and directs the reader's thinking toward as yet unresolved problems' rather than 'simply laying out a series of separate studies'.

Within the traditional framework adopted, the range and coverage in detail of the experimental work that is discussed leave little room for criticism. There is a wealth of excellent graphs and diagrams. In some places there are perhaps fewer tables than one might like to see. Partly on this account, the danger is if anything that the student who seeks his information here may be lured away from the 'experimental literature' altogether, with adverse effect on his own powers of appraisal and criticism of original data. The author, too, largely fulfils his claim that he has tried to write so as to indicate unsolved problems to the reader. Necessarily his suggestions are based upon his own conceptions; to this extent he must at once arouse reservations and provoke controversy, if not outright disagreement; but there can be no doubt that this form of presentation makes for a more stimulating text-book than would any mere classified directory of experimental psychology.

It is when one looks for 'an interpretative framework' in a book about 'method and theory in experimental psychology' that one finds less than might have been expected. The book does not attempt to approach the whole of experimental psychology consistently from one theoretical standpoint. Though foreshadowed in a discussion of constancy in perception, it is not until half-way through the book that the author's main theoretical contribution is encountered; at the end of a chapter devoted to theories of learning he discusses his 'mediation hypothesis' in attempted reconciliation of the field theoretical and behaviouristic approaches to learning. It would be inappropriate to embark upon a detailed criticism, but some comments are in order in examining the approach to theory and theories that the book adopts. The hypothesis is not put forward following any methodological analysis of other theories. It can be held that there is a failure to appreciate radical differences between them at this level and, to this extent, that the mediation hypothesis provides an eclectic basis for the psychologist whose object is to carry out experiments, rather than that it affords any substantial theoretical reconciliation. Much the same commentary is to be passed upon the partial discussions of other problems, met with here and there through the book, such as that of similarity and generalization, and those that arise in relating physiological, behavioural and phenomenal evidence.

In reply the author could justly observe that he would never have written a text-book at all if he had set himself a different kind of task, and could point to what he has done. What he has done is indicated by such a chapter heading as 'certain controversial issues in learning theory'. A careful statement of the different current positions on particular issues is given, selected experimental evidence is presented and evaluated, and places where there is a lack of evidence are indicated. There are similar discussions of, for example, theories of thermal sensitivity and of the positions of Hebb and Gestalt psychology on perceptual organization. The level of theoretical evaluation is perhaps brought out as clearly as anywhere in the chapter on 'serial and transfer phenomena'. Here the author develops, as he has less fully elsewhere, a systematic approach to much seemingly contradictory evidence, the discussion culminating in a theory at the level of empirical generalization of a high order. Crucial to the theory is, however, a quantitative conception of *similarity*. The issue of whether, and if so how, a theory can usefully employ such a concept is not pursued, even though in another chapter the possibilities of circular reasoning that inhere in its use have been briefly indicated.

'Methods and theories in experimental psychology' might therefore have been a more appropriate title. But again, these terms are apt to be all things to all men and perhaps particularly to psychologists. It remains to welcome a refreshing text-book which is from its standpoint a stimulating account of a large part of what is current in experimental psychology.

ARTHUR SUMMERFIELD

Psychometric Methods (a new second edition). By J. P. Guilford. New York: McGraw Hill Publications Inc. 1954. Pp. ix+597. 61*s*.

Many readers will be familiar with the first edition of *Psychometric Methods* which appeared in 1936. In the past, students of psychology have often turned to Guilford for statistical guidance. The new edition of *Psychometric Methods* is both longer and more selective than the earlier one. The comprehensive treatment of such statistical topics as frequency distributions, the normal curve, correlation and regression has been omitted since these have been included in the second edition of Guilford's *Fundamental Statistics in Psychology and Education*, published in 1950. This step has enabled the treatment of psychophysical methods to be extended and comparatively recent developments as well as the traditional methods are included. A mere catalogue of the contents of the sixteen chapters would convey little indication of the scope of the present volume. However, a broad summary of the contents will be useful. The first three chapters are essentially introductory, ranging from an elementary discussion of psychological measurement and psychophysical theory to a mathematical introduction intended to supply the minimum of necessary background knowledge for the benefit of those without formal training in mathematics. The next three chapters (4–6) deal, in turn, with the classical psychophysical methods. Chapters 7–10 deal with progressively more recent developments in psychophysical methods ranging from the Method of Pair Comparisons to various proposals for 'scaling' data based on human judgement. A lengthy chapter (11) is devoted to the Rating Scale as a psychological instrument. Whereas this forms an excellent summary of available facts, little seems to be added to what was already known and compiled. A most important chapter (12) is devoted to Principles of Judgement. In this Guilford summarizes the main factors influencing human judgement, and in doing so collects together for the first time the results of many recent psychophysical studies. Some of these were first published as late as 1952. From the point of view of the interested specialist this chapter will be one of the high-lights of the present volume. Guilford now turns his attention to a different set of topics. The next three chapters (13–15) are devoted to various psychometric problems in the field of mental testing. Such questions as item analysis, homogeneity and heterogeneity of tests, speed versus power tests, reliability and validity are discussed. The chapter on Test Development includes a brief reference to bias in the response to test items and methods which may be employed to minimize such bias are outlined. Attitude-scale construction occupies a few pages at the end of this chapter. The last chapter presents an admirable exposition of the theory and practice of Factor Analysis with emphasis on the American point of view. It is no mean accomplishment to have stated so much with clarity in so short a space. One may not agree with current American practice in Factor Analysis, but since it represents one important thread in contemporary psychology it is desirable for a concise yet effective statement of the procedures in use to be available. This chapter provides such a review, including worked examples.

The typography and layout of this volume are excellent, and remarkably few errors have been detected. Each chapter includes a lengthy bibliography and a few problems. At the end of the volume are appendices A–M which consist of a variety of statistical and mathematical tables in common use.

A critical review of *Psychometric Methods* would require more space than is available here, and more time for consideration than can be devoted to this review. However, a rapid appraisal is desirable. The student who wishes to obtain guidance on how to carry out one of the many specific techniques dealt with in this volume may consult it to his advantage. However, the more sophisticated reader will often retain mental reservations concerning the assumptions which have to be made in order to proceed with the more elaborate techniques. In a volume this size more space might profitably have been devoted to the qualitative implications of the assumptions which particular techniques force upon the investigator, and of the precise meaning of their end-products. It is quite clear from short interpolated paragraphs that Prof. Guilford is well aware of the necessity of stating clearly what assumptions can legitimately be made. However, he repeatedly throws upon the reader the responsibility for a decision on this point in applying the techniques outlined to practical problems. The beginner, for whom the mathematical introduction is intended, will not be in a position to decide such questions for himself. Furthermore, the sections on psychological scaling, on the psychometrics of mental testing and on Factor Analysis seem to be more advanced than the non-mathematical reader can approach in an intelligent and critical manner. Readers in the United Kingdom may regret the scanty reference to relevant work originating in this country. However, much of this work could not have been taken into account without reducing the reference to American work which is perhaps less accessible to many readers.

It might well be argued that we have here not one but the makings of more than one volume. Although one might just agree on the relevance of Factor Analysis to the main theme the three chapters on psycho-

metric problems of test development seem to belong to a different 'dimension' from the rest of the book. Individual sections will be of value to the general reader, and will certainly stimulate the expert in the field concerned. *Psychometric Methods* is probably best regarded as a reference book which should be included on the shelves of the psychological section of all libraries. Indeed the present reviewer had sufficient confidence in the importance of what Prof. Guilford writes to recommend that this new edition be purchased without having first scrutinized it in detail. This action has been fully justified by the book itself, even if one does not share all opinions expressed or the assumptions made.

ROLAND HARPER

Motivation and Personality. By A. H. Maslow. New York: Harper. 1954. Pp. xiv + 411. $4.50.

We have here what is really a collection of essays, most of which have appeared in earlier versions as journal articles. The author champions his socially conscientious, 'holistic-dynamic' conception of science in general and psychology in particular, and then proceeds to outline his well-known hierarchical theory of motivation. The theory distinguishes five levels to which the various motives of human beings can be assigned: physiological needs, safety needs, belongingness and love needs, esteem needs, and needs for self-actualization. The needs belonging to each of these levels emerge when those of the preceding level have been gratified. Maslow identifies progress along this sequence with an increasing degree of mental health. This leads him to recommend gratification of needs as far as possible and to feel that satisfaction of thwarted motives plays a greater part in psychotherapy than is usually acknowledged. The later chapters develop the implications of this theory for psychological health and ill-health, psychotherapy and cognitive processes.

Very little of the book is concerned with specifying the predictions that would enable one to test the theory or to see exactly how it differs from alternative theories. The evidence is mostly anecdotal, including case-history-anecdotal, or derived from clinical impressions. There are cursory appeals to controlled studies whose precise implications would bear prolonged discussion. Some interesting passages are sandwiched between generous slices of persuasive language, exhorting the reader to change to the author's preferred brands of classification and imagery.

Maslow's 'holistic-dynamic' principles seem to be adequately summarized by the assertion that behaviour is strongly influenced by motivational factors, and the assertion that behaviour is a non-linear function of a large number of variables. One would have to go back a long way to find a reputable psychologist who was unaware of these truths. But Maslow feels impelled to spend an astonishing amount of time berating a troop of straw men on dead horses, who, one gathers, make up the bulk of contemporary psychologists. These benighted individuals adhere to an 'atomistic-mechanistic-static' point of view: they appear to think that every response has a single external cause and is quite isolated from every other response, to explain learning by fortuitous combinations of external events which have no relevance to biological needs, and to have a liking for animal experiments without having realized that subhuman and human animals are different. He admits in the preface that these polemics, originally written about 1940, are now 'somewhat dated'. One would have thought that even such pioneers as Pavlov and Thorndike had written enough to absolve them from these errors, while their various successors have been proclaiming their avoidance of them for a quarter of a century. It is noteworthy that Maslow, who at other times provides copious references, gives none at all when attacking 'behaviourism' and that the abundant works of neo-behaviourists on the topics he handles are all but completely absent from his bibliography.

Maslow begins the book with a protest against an 'excluding, critical' view of science and in favour of a 'creative and constructive' one. It has evidently not struck him that people do not have to wait for scientists to supply them with ideas, least of all about human beings. But what has always been badly needed is a means of distinguishing the true from the false among the numerous hypotheses that are on the market. Some readers will fail to see why a strict application of the canons of scientific inquiry should be the same as a narrow one. To them the word 'science' will be inseparable from the ideals of precision and rigour and from a humble and critical attitude even to the assumptions one finds most acceptable. They will prefer unsolicited testimonials from clear-cut empirical data to the competitive skills of the copywriter, and the history of other sciences will convince them that these tastes are not inimical either to creative imagination or to social significance. Such readers will find the role assumed by Maslow too near that of the sage or the publicist for their liking.

D. E. BERLYNE

Learning Theory, Personality Theory and Clinical Research. The Kentucky Symposium. London: Chapman and Hall. 1954. Pp. 164. 28*s*.

Psychological literature has, of late, been increasingly encumbered by the publication of symposia in book form. More often than not, these lose in coherence what they gain in richness, the final product being a jumble of miscellaneous discourses rather than a unified treatise. So with this volume. It comprises eleven lectures delivered at a symposium designed to encourage a clearer integration of the three branches of psychology named in the title. The lecturers were: D. K. Adams, R. B. Ammons, J. M. Butler, R. B. Cattell, H. F. Harlow, N. R. F. Maier, O. H. Mowrer, D. Snygg, K. W. Spence, D. D. Wickens and J. R. Wittenborn. Taken individually, the lectures are able, stimulating, and informative. They reflect the characteristic research interests of their authors and give little cause for complaint. Collectively, however, they are less satisfying. There is but slight inter-relation between them and the co-ordinating pencil of an editor is nowhere evident. It is perplexing to speculate on the type of reader at which this work is aimed. For the layman, the lectures are too specialized. For the research worker, they are insufficiently detailed. And the professional psychologist can turn to recent issues of the *Psychological Review* and find there a collection of papers very similar to and as diverse as those published here. It is never unrewarding to read what such distinguished writers as the above have to say. But this fact, in itself, seems to lend slender justification to the production of a book which cannot be said to fulfil the ambitious and laudable intention of those who organized the symposium.

I. M. L. HUNTER

Studies in Schizophrenia: a multidisciplinary Approach to Mind-Brain Relationships. Reported by Robert G. Heath. Harvard University Press; London: Geoffrey Cumberlege. 1954. Pp. xiv + 619. $8.50.

Schizophrenia has been investigated by many different techniques in many disciplines. Thus, at the beginning of this monograph Dr Heath, the chairman of this multidisciplined group at Tulane University, has at his disposal a number of ideas from psychiatry, psychology, physiology, biochemistry, neurology and neurosurgery, Under the title 'Theoretical Concepts', a unified approach is sought to provide a framework for the study for schizophrenia. Despite the fact that the work starts off with a discussion of theory, the empirical nature of the technique eventually devised cannot be disguised. This by itself is not criticism of the methods used in studying the disease of schizophrenia, or the therapeutic approaches suggested by these methods. The empirical is the contemporary rule in these matters. However, a more formal and critical review would appear necessary before the claims of the subtitle of the work could be upheld.

Briefly the formulation is that patterns of overactivity and underactivity, of motor, psychological and chemical homeostatic mechanisms are attributable to excitation of hypothetical circuits in the brain. These facilitatory and inhibitory circuits exist as cortical to subcortical connexions.

From electrophysiological experimentation and animal investigations, attention has been drawn to the importance of the dynamic interaction between cortex and the septal region of the forebrain. In schizophrenia, it is argued, an impairment of facilitatory excitation is involved. Therefore stimulation of the relevant circuit at the level of the septal region will produce beneficial changes in those patients suffering from schizophrenia. This is tried on a group of twenty patients, mainly chronic schizophrenics.

The technique of stimulation is by surgical or stereotaxic implantation of electrodes in the relevant subcortical structures and by bringing the leads out to a plug attached to the skull. Thus patients can be treated by subcortical electrical stimulation at some time after the neurosurgical procedures. Some degree of therapeutic success is noted, but fortunately the monograph is content to point out therapeutic possibilities only and not to advance therapeutic claims.

All this, of course, is by and for electrophysiologists, neurosurgeons and the like. The course of the investigation is dictated by their interests and techniques. The psychologist and psychiatrist play a secondary role in trying to describe modes of changes and directions of clinical improvement. In this respect the psychologists involved have profited by their experience in the Columbia–Greystones project and have contented themselves with a narrower range of investigation—namely to find a battery of tests which reflects the changes produced in patients by this stimulation technique. The battery in question included the Eisenson Tests of aphasia, a shortened Wechsler-Bellevue, Porteus Mazes, Benton Visual Retention Rest, Rorschach, T.A.T., a Behaviour Rating Scale, and a group of psychomotor tests which seem highly appropriate to this kind of work, but which disappointingly are not discussed or described.

Dr H. E. King, their author, however, apparently intends shortly to communicate a description of these tests and their uses.

While the psychological contribution by itself does not add much to our little knowledge in this field the monograph as a whole represents one extension of present physical methods of treatment which may in the future concern clinical psychologists.

R. M. MOWBRAY

Fundamentals of Psychoanalytic Technique. By T. Braatøy. London: Chapman and Hall. 1954. Pp. ix+404. 48*s*.

In this book the late Trygve Braatøy gives an account of the methods of analytical psychotherapy which he developed as a result of a very large psychiatric experience in his native Norway, in France, Germany and the United States.

Braatøy was much influenced by the developments of Freud's ideas by Otto Fenichel and Wilhelm Reich before the second World War. Because of his wide interests in clinical medicine, physical medicine and medical ethics as well as psychiatry he became increasingly worried by the tendency of psychoanalysis to become divorced from other forms of medical treatment. Gradually he evolved his own principles and methods of analysis and in this book he describes them.

Because of the demands analytical treatment makes on the energy and time of the patient and the therapist and because it may be dangerous psychoanalysis is restricted to a relatively small number of patients who should be carefully selected. The author gives his criteria for deciding that analysis is indicated and stresses that account must be taken of the qualities of the analyst as well as those of the patient.

The major part of the book consists of detailed descriptions of the author's principles and methods of treatment, the difficulties encountered and the means to avoid them. He presents numerous case-histories and it is refreshing and very instructive that he discusses his failures, as well as his successes.

The book is practical in its approach. The position of the patient on the couch and the need for him to be relaxed are discussed in much greater detail than in Freud's concept of personality structure; techniques, not ideals, are its concern. At the same time the author justifies his methods with reference to a wide range of analytical and non-analytical literature. His common sense and insistence that accurate observation and diagnosis are essential for satisfactory psychotherapy leads to some penetrating criticism of a number of latter-day Freudians. But his criticism is of their practice not of their principles, and it is constructive, not destructive. In particular his review of the methods of Reich and Fenichel is most illuminating. The author's concern with technique rather than theory is illustrated by the fact that there is no reference to the works of Jung in the bibliography. It might well have been called 'Psychoanalysis in practice; a personal interpretation of Freudian principles'.

Braatøy wrote enthusiastically and the book tends to be diffuse and repetitious in parts. The English is usually adequate to make clear the meaning of the author but is never elegant. Norwegianisms abound. Many of the very interesting footnotes could easily have been incorporated in the text. As it is they are so numerous that they prove distracting. Since the book is written in English it would have been a convenience if translated German works had been quoted in their English versions in the bibliography.

The author states that he assumes his readers have a considerable familiarity with the major works of Freud. At least some experience of the practical problems of psychotherapy is required for the book to be fully appreciated. It is in no sense an elementary text-book and the student would probably be puzzled by it rather than enlightened. It is a fascinating and valuable account of the practical problems encountered by one courageous and individualistic psychiatrist in attempting to reconcile the principles and methods of psychoanalysis to those of other forms of psychiatric and medical treatment. The fact that the verdict must be that the reconciliation he accomplished remains incomplete makes one the sorrier that this is the last book we shall have from this stimulating psychiatrist.

T. T. S. INGRAM

The Collected Papers of Otto Fenichel. First Series. Collected and edited by Hanna Fenichel and David Rapaport. London: Routledge and Kegan Paul Ltd. 1954. Pp. xiv+408. 35*s*. net.

The Collected Papers of Otto Fenichel are being published in two volumes; and the present volume represents the First Series, consisting of papers written between 1922 and 1936. Fenichel, who died in 1946, was a psychoanalytical writer of unusual distinction; and his book *The Psychoanalytic Theory of the*

Neuroses is a classic work. An as expositor of Freudian theory he is unrivalled; and indeed his grasp of essentials and capacity for systematization, combined with his terse, lucid and forthright style, often render his expositions of classic psychoanalytical theory considerably more effective than those of Freud himself. At the same time, his sincere adherence to basic Freudian principles (and his obvious sincerity cannot fail to impress the reader) does not in any way impair the exercise of his critical faculty; and he does not hesitate to exercise this faculty, with a respectful cogency which sets a standard for scientific argument, whenever he feels criticism of any specific aspect of Freud's thought to be justified. Thus his criticism of Freud's theory of the death instinct in a paper devoted to this subject is perhaps the most cogent criticism of this theory on the basis of Freud's own principles that has ever appeared in print. The two papers entitled 'Psychoanalytic Method' and 'Concerning the Theory of Psychoanalytic Technique' contain masterly expositions of the principles underlying psychoanalytical therapy in its classic form. Two other important theoretical papers are those entitled 'Identification' and 'Defense [*sic*] against Anxiety, Particularly by Libidinization'. The remaining papers are predominantly clinical; but many contain theoretical formulations of considerable interest and importance. The depth of Fenichel's clinical insight cannot fail to command the respect of even the most experienced psychoanalyst. His elaborate interpretations of some of the clinical material which he describes may well mystify the reader with no special psychoanalytical training; but to the trained psychoanalyst they will present themselves as evidence of an unusual grasp of the complex currents of unconscious motivation on the part of Fenichel. The book as a whole is one which is bound to appeal to all who are interested in psychoanalysis, and will stir in all such the hope that there will be no undue delay in the appearance of the Second Series of Fenichel's papers.

W. RONALD D. FAIRBAIRN

The Development of Personality. By C. G. Jung. London: Routledge and Kegan Paul Ltd. 1954. Pp. vii+235. 25*s.* net.

This is volume XVII of *The Collected Works of C. G. Jung*, and the fourth volume of the series to be published. It comprises eight papers, of which all except the last two ('The Development of the Personality' and 'Marriage as a Psychological Relationship') are devoted to child psychology and education, and are addressed primarily to educationists. According to Jung, the psychological development of the child consists essentially in the development of consciousness out of the original unconscious, which is partly collective and partly due to an 'identity' of the child's psychic state with the unconscious of his parents. Differentiation of the ego is regarded as characteristic of this process of development in so far as its aim is successfully achieved; and the first six papers are devoted largely to a description of the process in question. From a practical standpoint, the most important bearing of Jung's theme arises out of his view that the psychological problems of the child are chiefly a reflexion of the unconscious problems of his parents, and secondarily those of his educators as parent-figures. To the reader who has found light thrown upon the development of the child's ego by the views of Freud and Melanie Klein, doubtless the most remarkable features of these papers will be Jung's soft-pedalling of infantile sexuality and his minimizing of the significance of aggression; but the reader will also perhaps note with surprise the correspondence between Jung's theory of the development of consciousness out of the unconscious and Freud's theory of the development of the ego out of that part of the id which makes contact with outer reality.

The last paper, on 'Marriage as a Psychological Relationship', is devoted essentially to a favourite theme of Jung's later writings: viz. (as indicated in an Editorial Note) his conception of personality as 'an adult ideal whose conscious realization through individuation is the aim of human development in the second half of life'; and, in accordance with this conception, Jung expounds the view that it is only as the result of successful individuation achieved in middle life that marriage can acquire the status of a true personal relationship. The second last paper, on 'The Development of Personality', is essentially a description of that sense of vocation which has characterized the leading figures in human history; and for many readers this will doubtless prove the most interesting paper in the book. The scientific psychologist may be disappointed to find that Jung's illuminating description of what is involved in the sense of vocation is not matched by an explanation of its psychological origins; but he may also console himself with the reflexion that, when reductive explanations of such sublime phenomena are attempted (e.g. along Freudian lines), much of what is to be explained is commonly lost in the very process of explanation.

W. RONALD D. FAIRBAIRN

The Prediction of Performance in Clinical Psychology. By E. Lowell Kelly and Donald W. Fiske. Published by the University of Michigan Press, Ann Arbor. 1951. Pp. xv + 311.

This is an account of a large-scale research into the prediction of performance in clinical psychology. The authors tell us that at least ten Ph.D. theses went into this piece of work which stretched over some five years. The contributors to the project included many well-known names in the field of mental testing and personality assessment.

Almost every known method was used to predict aptitude for clinical psychology. From a study of this mass of material, the authors state that the Miller Analogies test was more significant than the Thurstone's P.M.A. and that less weight could be given to the M.M.P.I., the Allport-Vernon Scale of Values and to most of the Guilford-Martin Scale. Some sections of the Strong Interest Blank were useful. The user's confidence in the interview is an 'extremely fallible index of its actual validity in a specific situation'.

The various members of the research team were also equally thorough in their search for suitable criteria to take as measures of achievement in clinical psychology. Finally, they arrived at thirteen variables which include academic performance, clinical psychology and general psychology, diagnostic competence, individual therapy, group therapy, research, supervision, interpersonal relations, integrity, clinical competence, preference for hiring, and liking. The criteria were factor analysed and as a result three factors were identified which seem to characterize clinical performance. They were an *intellectual factor*, effectiveness in close *interpersonal relations* and *personal integrity*.

The authors tackle the problem of validity piecemeal after indicating the general pattern of the prediction problem (p. 124). First the validity of university predictions over a period of three years is considered. These are generally the highest of all the correlations obtained. Then the validity of the objective tests are set out and the intellectual ability tests are taken first. The Miller-Analogies test comes out very well with correlations of the order of 0·35 and 0·47 for various parts of the multiple criterion. The *R* and *M* subtests of the P.M.A. test are most suggestive of prediction and in the Guilford-Martin battery, thinking extraversion, freedom from cycloid tendencies, and social extraversion yield the highest correlations with related parts of the criterion. These include diagnostic competence, individual psychotherapy, interpersonal relations, integrity and preference for hiring. Research competence, academic performance, clinical competence, preference for hiring and the marks in clinical and general psychology are predicted well by the group I of the Strong vocations, particularly those referring to the artist, psychologist and architect. In the group II occupations, the mathematician and physicist responses correlate with academic performance and research competence. A review of the higher correlations obtained from these different objective tests is given on page 158. The validities of the personality assessment ratings are, on the whole, quite high, ranging up to 0·5. One interesting result shows that intervening personal therapy tends to reduce the validity of personality assessments.

It is a pity that the authors did not have recourse to multivariate analyses in order to obtain the maximum prediction from a selected number of predictors. If it were objected that such a task would be extremely lengthy, with so many components in the criterion and in the predicting side of the battery, one could suggest that use might be made of the discriminating function. The book is not easy to read, particularly for the English reader. It presents a trying mixture of heavy passages of technical jargon interspersed with tables of extremely condensed material. Lastly, it is, perhaps, somewhat chastening to find that certain tendencies already discovered in less ambitious programmes were merely confirmed.

However, the work as a whole should be studied by all engaged on the prediction of success in a complex professional field. The 'general outcomes' listed on page 197 are valuable and follow from the data. Some of the methods described and the results obtained are very pertinent to current studies being carried out in this country on the selection of medical students and teachers and for other professional occupations. E. A. PEEL

Deprived Children. The Mersham Experiment. A Social and Clinical Study. By Hilda Lewis, M.D., M.R.C.P., with a foreword by Sir Arthur Ellis and Dr C. P. Blacker. Published for the Nuffield Foundation. London: Oxford University Press. 1954. Pp. xiii + 163. 9*s*. 6*d*. net.

This book is a scientific assessment and evaluation of the work of the Mersham Reception Centre. This Centre arose as a pilot experiment on the instigation of Miss Leila Rendel, the Director of the Caldecott Community, to the Nuffield Foundation. Miss Rendel was 'gravely disturbed at the haphazard way in

which deprived children needing care were distributed to children's homes, foster homes or approved schools without any real study of their individual needs or difficulties'. As a result the experiment, which was initiated in 1944, was sponsored by the Caldecott Community whose 'solid background of achievement and experience in the case of deprived children was available for this first study of the problems involved in a Reception Centre'.

A description of the Reception Centre and its work is given. A high proportion of the neurotic and maladjusted children were referred to the Centre. One-twelfth of the children received were recommended for return to their own homes, one-fifth to foster-homes and slightly more than one-third to children's homes. 76% of the children were placed according to the recommendation made. Certain practical difficulties arose because facilities in the county of Kent were restricted, although these became easier after the passing of the 1948 Children Act.

A fairly detailed account is given of the 500 Kent children admitted between October 1947 and July 1950, 277 (55%) of whom were boys. Information was obtained on their family backgrounds in terms of social class, size, income and living conditions, and a careful study was made of the 363 families, information being available concerning 264 fathers and 328 mothers. There is evidence to show that the families were larger than the average for the county, that their incomes and occupational stability were low and that a fair proportion (30%) were living in over-crowded conditions (comparable figure for the general population of Kent was 1·09%). Relationship between the parents was estimated as definitely bad in 45% of the families. Medical disabilities especially mental disabilities, both defects of intellect and personality, were common (99 fathers and 180 mothers).

A quarter of the children admitted to the Centre were normal in behaviour and general mental condition. 4% were classed as psychotic or psychopathic, 32% as delinquent in some degree and 18% showed severe neurotic symptoms. There was a slight preponderance of dull children, but 47% were of average intelligence. Attainment in reading and arithmetic was poorer than age or intelligence warranted. 21% were in poor physical condition. It is noteworthy that the two most marked psychological characteristics of the children were excessive demands for attention and lack of emotional warmth in their reactions.

An attempt was made to classify the general behaviour of the children according to the patterns described by Hewitt and Jenkins, viz. unsocialized aggression, socialized delinquency, over-inhibited neurosis and normal behaviour. This is of interest to the psychiatric reader, but it is doubtful whether such a classification is of great usefulness.

The analysis of the data obtained by so qualified a team of observers as psychiatrist, psychologist and psychiatric social worker is of great value. It is important to find evidence to show that five background features had a statistically significant association with patterns of disturbed behaviour, i.e. neurotic or psychopathic mother, a mother lacking in affection for the child, prolonged stay in public care and illegitimacy. It is valuable also to find the concept that early separation from the mother invariably produces serious emotional effects carefully examined. There was no evidence that delinquency or incapacity for affectionate relationships was significantly more frequent in the separated children. Children who had suffered lasting separation from their mothers, which began before the age of two, tended to be more disturbed in behaviour than those who had not been exposed to lasting separation at any age however. But the report emphasizes the importance of considering the quality of the child's relation to his mother before deprivation and his experiences with mother-substitutes while separated. It also stresses the emotional hazards and harms children may suffer from bad mothers and indifferent mother-substitutes and the need for more sober research into this whole matter.

There were 66 'problem families', characterized by gross neglect and squalor, represented in the sample. Although many of the parents were handicapped in personality or intelligence they were affectionate to their children, who though of lower intelligence, were freer from delinquency or neurosis than the rest of the Reception Centre children. This supports other findings in regard to this group.

The follow-up study which was made on 240 children after an interval of 2 years or more and included a special intensive inquiry on 100 children, revealed that the proportion of children in good psychological and social condition had more than doubled. Children who had been placed in accordance with the recommendations of the Centre had done somewhat better than those who had not been placed as recommended. On the whole, children placed in foster homes proved most satisfactory though not invariably so. Close contact with his own family proved greatly to the child's advantage in general. Good intelligence was also prognostically favourable.

Any serious reader of this book will be impressed with the care and scientific caution by which the work of the Reception Centre is evaluated. With what is described by Dr Blacker and Sir Arthur Ellis as the

utmost emotional neutrality, Dr Lewis has shown the very great value of the work of a well-run, adequately staffed Reception Centre in assisting in the difficult and delicate task of finding the appropriate placement for the particular child. Her findings are supported by the Childrens' Officer of Kent. The various alternative methods are examined in the book and their disadvantages demonstrated and the experience and evidence accumulated at the Mersham Centre in Dr Lewis's words 'compel the firm conviction that reception centres, though not ideal, are indispensable at the present time for dealing with a large proportion of deprived children', but that 'a bad reception centre is probably worse than none at all'.

This is a book which all workers in the field of Child Care cannot neglect to read.

AGATHA H. BOWLEY

Children's Humor. By MARTHA WOLFENSTEIN. Glencoe, Illinois: The Free Press. 1954. Pp. 224. Price $3.75.

This book contains examples and analyses of children's humour at different ages. The analyses are based on the theory of Freud as developed in his *Wit and its Relation to the Unconscious*. In addition to its manifest content, the joke, like the dream, has also a latent meaning which has undergone distortion and disguise. In the introduction, the author suggests that some readers may be inclined to feel that more is seen in the jokes than is there, with which statement the reviewer tends to agree.

From an early age, the child uses the joke to enable him to alleviate difficulties. He endeavours to transform painful experiences into enjoyable ones. The basic motive is to triumph over distress and to gain relief, if even momentarily, from frustration. The 5-year olds try to cope, through joking, with their envy of the bigness and power of adults. From the age of 6–11, the latency period, jokes tend to be identified with riddles, and the 'moron' who represents all that the child repudiates in himself, becomes the main protagonist. At adolescence, the riddle is replaced by the anecdote, and mimicry becomes a component of joke telling.

The five chapters in turn deal with Joking and Anxiety; Sex, Names and Double Meanings; Riddles and the Legend of the Moron; Development of the Joke Façade; Understanding the Joke. The contents of the chapters are interesting, and the analyses of the jokes throw light on certain aspects of childrens' development, although one wearies of the constant Freudian interpretation.

The book can be recommended as a useful contribution to an understanding of humour, as well as a careful study of a specialized field.

M. COLLINS

Psychology of Learning and Teaching. By H. W. BERNARD. New York: McGraw-Hill Publications Inc. 1954. Pp. ix+436. $5.00.

It is nowadays important to note how far the pupil is a product of social circumstance and the scope of educational psychology has widened accordingly. Dr Bernard thinks and writes in this tradition, stressing 'the shift in emphasis from physiological psychology to one oriented in social groupings'. So we have sections given over to 'the role of the social setting' and the teacher's 'dynamic influence' on the pupils' mental hygiene.

He underlines also the continuous nature of child development and 'growth as a pervasive factor in learning and teaching'. Finally, evaluation itself is seen as 'a continuous process rather than as a culminating act'. This is a view which will find sympathy in Britain, for we know only too well of the dangers of taking a single cross-section when we test progress and of the value of cumulative records.

The material in *Psychology of Learning and Teaching* is well presented, but is perhaps a little full and even wordy for a first text-book. Modern theories of learning, motivation and goal seeking are all invoked, but in spite of the useful summary provided at the end of each chapter, they are not sufficiently outlined for an inexpert reader. In general, the reader would be helped by a more explicit reference to theory. So on page 28 we read that 'not only are habits ways of acting, but, even perhaps more important they constitute a form of motive'. The reader's grasp of this important observation would have been enhanced had the author referred him explicitly to the concept of secondary reinforcement and related this to the other main features of Hull's theory. The main purpose of theory is to provide a comprehensive explanation and to give the teacher a frame of reference by which to relate and assemble the facts of his teaching experience.

The omission of these references to theory tends to give the book the appearance of a book of tips for learning and teaching. The British student might tend to find the scene and examples a little alien to him —as, for example, in the 'time schedule' and study tips given to student readers on pp. 26–27.

As a whole, however, the book is a useful description of many of the facts of learning, motivation and individual differences.

E. A. PEEL

A Theory of Social Control. By Richard T. LaPiere. New York, Toronto and London: McGraw-Hill Publications Inc. 1954. Pp. xi+568. 53*s.* 6*d.*

This is a book for sociologists. Its thesis is that man is a 'status seeking animal' and that this status is sought in relatively small and intimate groups which thus provide the medium in and through which social norms are operated and social control of behaviour is brought about. Thus, for the author, social control is one of the three major determinants of human behaviour, the other two being personality and situation.

It will be apparent that such an analysis depends for its cogency on clear, full, and somewhat arbitrary definitions of the factors involved. For LaPiere the distinguishing feature of social control is that it operates through the conferring or withholding of status, and the major part of the book is concerned with the techniques by which this is accomplished in groups and societies of different kinds.

For the social psychologist, whose primary interest is, or should be, measurement and experiment in the social sciences, the second part of the book, on the nature and operation of social control, is the most interesting and the least satisfactory. The author's procedure of stating his hypotheses and illustrating them anecdotally and by reference to work which is, in most cases, of an anthropological nature, is here least effective. For example, we would agree, in a purely common-sense way with a proposition of such high generality as: 'All other factors remaining equal, the control that is exercised by a group over an individual member is inverse to the size of the group', although, if pressed, we would probably have little difficulty in finding exceptions. But when this statement is translated by the author into a proposition of lower generality, namely: 'Generally speaking, a group of ten can exercise ten times as much effective control over an individual as can a group of one hundred, a hundred times as much as a group of one thousand, etc.', then we feel that if words mean anything at all, it is time to call a halt and insist that rigorous, measurable definitions of the variables involved and experimental evidence of their relation, be forthcoming.

It should not be thought that the book contains many unsupported lower generality statements. Most of those presented are indeed the more sophisticated 'high' generalizations. This is unfortunate, since the use of 'high' language tends to obscure the necessity for empirical validating evidence; or to put it more bluntly, one can get away with more in a 'high' language than in a 'lower' one. (There comes to mind the paradox that Hull's greatness lies in the fact that he was the first psychologist who could be proved wrong.)

This work contains many interesting hypotheses for psychologists, but until the author uses a 'lower' language, we must regretfully repeat that this is a book for sociologists.

DEREK PUGH

Sex and the Social Order. By Georgene H. Seward. Harmondsworth: Penguin Books Ltd. 1954. Pp. ix+275. 2*s.* 6*d.*

This book was originally published in 1946. It must be said at once that it is a good book and fully justifies its appearance as a Pelican. It is exceptionally clearly written and easy to read, as well as informative, and can be recommended to anyone. The chapters are short and compact, and there are name and subject indexes and a bibliography of seven hundred items. If the author shows signs of suffragette zeal in her pleas for greater equality between the sexes, the book is none the worse for that.

Of the sixteen chapters, one is introductory and one, entitled 'Dynamics of Sexual Behaviour', deals briefly with the glandular and nervous basis of sexual behaviour. Five chapters are devoted to sexual behaviour in the animal world, from fish to chimpanzees, and four chapters to sex values and roles in primitive and in western society. Four chapters review the stages of sexual development from infancy to old age, and some of the associated problems, especially but not exclusively the social aspects thereof. One chapter gives a rather sketchy discussion of sex differences in modern (American) society. The last chapter, headed 'Sex in Post-war Society' contains a vigorous statement of the author's plea, which has already been made explicitly or implicitly a number of times, for a reformulation of sex roles in the

interests of mental health and social efficiency. At times, the author is perhaps a little uncritical, and some topics which appear relevant have been omitted, for example, the problem of differential fertility and its possible effects.

One is left wishing that the author had dealt with her subject at much greater length. But a longer book would perhaps have proved less enjoyable reading.

D. GRAHAM

Character and Social Structure. By HANS GERTH and C. WRIGHT MILLS. London: Routledge and Kegan Paul. 1954. Pp. xxv+490. 30*s*.

In the preface to this book, the authors state that 'our general purpose is to study the personalities of men in connexion with types of social-historical structure. We wish to analyze conduct and character by understanding the motivations of men who occupy different positions within various social structures'. This promises well, but to at least one reader, this promise is not fulfilled, and the book is to this extent disappointing. Despite the subtitle 'The Psychology of Social Institutions', the book is much more of an essay in sociology than a contribution to psychology. In Part One, the authors distinguish between organism, psychic structure ('integration of feeling, sensation and impulse'), person ('man as a player of social roles') and character structure ('integration of the organism's psychic structure with the social roles of the person'). One might expect most of the book to be concerned with character structure as thus defined.

Part Two, entitled 'Character Structure' draws heavily on the ideas of G. H. Mead. There is an interesting chapter on the sociology of motivation and the attribution of motives. What strikes the reviewer most, however, is the absence of any proper attempt to link the analysis of character with learning theory.

But most of the book is concerned with analysis at the institutional level. Parts Three and Four, entitled 'Social Structure' and 'Dynamics' respectively, are uneven. The chapters on 'Symbol Spheres' and 'Stratification and Institutional Orders' are rewarding, those on 'Institutions and Persons', 'Institutional Orders and Social Controls' and 'Collective Behaviour' are less enlightening. Material is drawn from historical periods as far apart as Ancient China and the contemporary world. The overall result is somewhat disjointed, and one wonders whether it would not have been better to restrict the canvas in one way or another. In particular, one may question the advisability of trying to deal, even in a cursory way, with the characteristics of the principal world religions in less than four pages.

If this book had to be summed up in a single sentence, it would perhaps not be unjust to say that it has attempted too much at the institutional level, and achieved too little at the psychological level. This is, of course, a psychologist's judgement.

D. GRAHAM

The Psychology of Politics. By H. J. EYSENCK. London: Routledge and Kegan Paul. 1954. Pp. xvi+318. 25*s*.

When we come into contact with clear, original and forcefully expressed ideas, it is almost axiomatic that we say, perhaps somewhat wistfully, 'Now why didn't I think of that myself?' It sounds so easy, and Prof. Eysenck's conclusions have just this quality of deceptive simplicity. But what the man in the street tends to forget is that accurate simplicity, as opposed to naïve hit or miss, is hard work. The detailed experimental work and the tough conceptualizing which have enabled Prof. Eysenck to summarize his conclusions in a few lucid pages, compel respect and admiration.

This book, selective as science must be, does not pretend to describe all the historical, economic, and religious, etc., ideas which govern political affiliation. It is merely concerned to discover whether anything can be said from the psychological point of view; whether a psychological approach can reveal qualities held in common by persons of the same political outlook. Briefly, the conclusions are as follows.

Political attitudes are governed by two quite independent factors: the Radicalism–Conservative continuum (Prof. Eysenck refers to its as the R-factor); and the Tough–Tender minded dichotomy of William James (the T-factor). Of these two, the R-factor is social in origin, while the T-factor is personal. Thus fascists are tough-minded conservatives and communists tough-minded radicals. Liberals are tender-minded and in the middle of the Radicalism–Conservatism continuum. In addition, there is a relationship of tough-mindedness with extraversion, with aggression and dominance and with 'such concepts as rigidity, intolerance of ambiguity, narrow-mindedness and mental concreteness'. The tender-minded, by contrast, are introverts.

These formulations would not, of themselves, prove of essential value to psychology, however much light they might throw upon the social scene, if they were not linked with more general issues. In particular, Prof. Eysenck attempts to relate his findings to learning theory since attitudes, being acquired dispositions, closely resemble habits. Viewed in this context, R and T-factors are respectively the neural modifications involved in learning and in conditioning, the former implying a change of behaviour through the influence of reward and punishment (and hence social in origin); the latter implying changes of behaviour derived from the involuntary emotional responses evoked by association.

This is the meat of the book. But for the specialist, its value will also lie in the detailed account of experimental method and procedure, in the honest exposition of uncertainty, and in the commentaries made *en passant* on much other psychological work.

This admirable book may not make the pioneer impact of *The Authoritarian Personality* by Adorno and his colleagues, but there was an undertow of emotion in that remarkable work which aroused certain doubts. In some respects, Prof. Eysenck has made a more solid contribution to psychological theory. In particular, and this is always the test, he has opened up many new fields for future research.

ADAM CURLE

The University Teaching of Social Sciences: *Sociology, Social Psychology and Anthropology.* UNESCO. 1954. Pp. 525. 10*s*. 6*d*.

This is the first of five UNESCO reports comparing the teaching of the Social Sciences in different parts of the world. The International Sociological Association carried out a survey of the extent and organization of teaching and research in eight countries. A 'rapporteur' was appointed for each country (Donald MacRae here), and the reports were discussed by a committee. Prof. de Bie of Louvain gives an overall review of the results, and there are several shorter articles, including one by Prof. Nuttin on Social Psychology, as well as factual summaries for the different countries covered. This arrangement involves a lot of repetition and could be improved upon in the later reports.

It is no surprise to read that there is a tremendous gulf between America and other countries in the extent of social psychology teaching and research. Furthermore, the subject has a broader scope there, while there is greater development of standardized techniques of research, and more team-work. From what information is given the subject has much the same content in many other countries as it has here, and the same American books are used, save in Germany and parts of France, where the old philosophical approach may be found.

This is a useful little book, though the British material is not always accurate—partly through being four years old now. A number of recommendations are made—that teaching should be linked with research, that better training in research methods be given, and that close co-operation with experimental psychologists and sociologists should be maintained.

MICHAEL ARGYLE

The Lowenfeld Mosaic Test. By MARGARET LOWENFELD. London: Newman Neame Ltd. 1954. Pp. 300+144 colour plates (in separate container). 50*s*.

The Mosaic Test was devised by Dr Lowenfeld more than twenty years ago, but this is the first text-book which has appeared from her hand describing the test in detail.

The book is divided into two main sections. Part I contains a description of the test itself, and discusses the method of classification of the results. The test consists of 456 pieces composed of different geometrical shapes—squares, half-squares, diamonds, equilateral triangles and scalenes—each coloured in one of six colours, red, blue, yellow, black, green and white. The individual is asked to make anything he likes on the tray provided, and also given instructions to use as many or as few pieces as he wishes, and any of the shapes or colours. The shapes and colours used are significant, as well as the type of design produced, and its position on the tray. Dr Lowenfeld describes in detail the different types of designs, and their classification which is exceedingly complex. Constant reference is made to the 144 coloured reproductions of designs which accompany the book.

Part II discusses the use of the test with different groups, and illustrates how the test is interpreted. A separate chapter is devoted to each group: Children, Individuals with subnormal intelligence, The Normal Personality, Neurosis, Mental Disorders (written by Dr Ellenberger), Cultural Problems.

The interpretation of any design depends on six aspects: (*a*) the manipulation of the material; (*b*) the relation between the pieces used and the space of the tray; (*c*) the form of the resulting design; (*d*) the

handling of colour; (*e*) the content; (*f*) the comments of the subject and his attitude to the test, and to what he has made. The author claims that the form of the Design represents a kind of self-portrait.

Some of the interpretations are extraordinarily interesting such as those of two sisters whose designs show striking similarity. The chapter on children's designs, tabulating the changes in the form of the design with age, is of value from the point of view of children's development. The designs indicating neurosis and mental disorders show features which differentiate them from the normal. Dr Ellenberger, who contributes the chapter on the use of the L.M.T. in the study of mental disorders, confirms Wertham's conclusions that the L.M.T. is an excellent and invaluable instrument for use in the differential diagnosis of diseases of the brain from mental diseases. The test, according to the same writer, has also yielded significant results with schizophrenics and can be used when other tests have failed. Dr Lowenfeld herself claims that the designs produced may detect nervous breakdown a considerable time before it actually occurs.

The different facets of the test seem unlimited, and if one can judge from the contents of the book, the test is being extensively tried out in different fields, and used in different countries. One considerable disadvantage from which it suffers is the difficulty of making a permanent record of the results. Apart from colour photography, white paper has to be placed on the tray beforehand, a tracing made before the pieces are removed, and finally the design has to be coloured—a time-consuming process.

A second disadvantage of the test, as it appears to the reviewer, is that the interpretation of the designs still seems to be largely subjective. As is stated in the book, to classify and to interpret the designs is like learning a new language, and the only reliable method of learning is said to be personal experience and direct personal teaching. 'It must also be granted that for the full use of the test and an appreciation of its subtler aspects, a certain special quality is needed in the investigator, one akin to the quality in the musician that makes elements of form and rhythm obvious to him that, to the unmusical, are imperceptible' (p. 48). If this is so, then the use of the test is certainly limited, and this unfortunately is the impression one receives from persual of the text.

While the book makes interesting reading, one would have liked the contents of some of the chapters to have been arranged more systematically.

M. COLLINS

The Thematic Apperception Test and the Children's Apperception Test in clinical use. By Leopold Bellak. New York: Grune and Stratton. 1954. Pp. x+282. $6.75.

As Dr Bellak points out (p. 147) TAT was originally developed by Henry Murray in an academic setting; not only the material but the general approach bears the mark of this origin, and it has been Bellak's aim to develop a technique of analysis and interpretation rather more fully appropriate to clinical use. Repeated reference is made to the need to accumulate norms, including separate study of different regional and cultural subsamples, but the author is careful to add that such norms are needed 'not to substitute for dynamic diagnosis, but to serve as a reliable background for it'. Bellak's major contribution in the earlier (and slightly longer) portion of the book, which deals with TAT proper, is the presentation of his own system of analysis, which, in so far as it is based on the use of a number of 'scoring categories' concerned more with content than with form, most closely resembles Wyatt's system (1947, *J. Psychol.* XXIV, 319–30), and stands in contrast to Rapaport's and to some extent to Murray's own, although something rather like need-press analysis is to be found in Bellak's listing of the 'main needs of the hero', together with 'conception of environment', 'parental (and other) figures', 'significant conflicts', 'nature of anxieties', etc. All this accurately reflects the author's clinical purpose; a more rigorous methodological innovation, reminiscent of Tomkins's approach, is his recommendation that the 'Main Theme' be progressively restated on various levels (up to five) of generalization or abstraction. Scrutiny of the data thus assembled, it is claimed, usually enables one to recognize a repetitive pattern in the subject's responses, and to proceed thence to a final report, which may or may not offer a diagnosis. Since he does not regard TAT as primarily a diagnostic instrument, Bellak recommends the use of some such formula as 'The data represented in the TAT are consistent with a diagnosis of...'. Such diagnosis, he says, should never be made on test evidence alone, without additional information provided by a clinical interview.

The second part of the book is devoted to a description of the Children's Apperception Test (CAT) and a Supplement to it (CAT–S). Based on a suggestion by Ernst Kris, the material consists essentially of animals mostly (though not always) in anthropomorphic guise. The CAT and the Supplement each consist of ten pictures, nearly all depicting or including 'animal children'. The situations in the CAT proper are regarded as more 'universal'; those in the CAT–S relate to more specific themes, such as temporary

disability. In addition, the CAT–S pictures are better suited to more informal administration; they are on curvilinear cut-out shapes which ask to be handled. Most users of TAT proper will agree that even Murray's 'children's' cards are not entirely satisfactory for use with children, and Bellak makes out a fairly convincing case for the substitution of animal for human figures. The age range recommended is three to ten, and the material is claimed to be relatively culture-free. This latter seems to this reviewer to be rather a questionable assumption: not only the humanized animals but some of the family-life situations are unmistakably American. Accordingly, it comes as rather a surprise to read in the Appendix on 'Publications and Work in Progress', that extensive work on the CAT is going on in India, France, the French colonies and elsewhere. One also gets the impression that too close a preoccupation with some of the stock 'psychoanalytic' and allied themes (e.g. primal scene, castration fantasy, the body-image, etc.) has been allowed to divert attention from the possibly more superficial but nevertheless pressing problems which form the staple of child guidance work, at least in this country.

It has been necessary in this review to concentrate on the descriptive side of Bellak's work, but the book also contains a good deal of valuable and thought-provoking general discussion of projective theory and practice. It is eminently a readable book, the author's style being both succinct and at times pleasantly colloquial.

BORIS SEMEONOFF

Sentence Completion. By J. Q. Holsopple and Florence R. Miale. Springfield, Ill. U.S.A.: Charles C. Thomas; Oxford, England: Blackwood Scientific Publications Ltd. 1954. Pp. xiii+177. 40*s.*

Subtitled 'A Projective method for the study of personality' this book presents a collection of 73 'openings', from which, it is claimed, the clinician may hope 'to obtain material from which [to] draw valid inferences concerning unconscious and semiconscious desires, motives, conflicts and systems of personality organization'. How this may be done is described almost exclusively by reference to illustrative records, which occupy about three-quarters of the book. Only one case, however, is interpreted at all fully; this subject's Rorschach record is appended, verbatim, but entirely without comment, except to call attention to 'the clear consistency which may be found when sentence completion and Rorschach materials are compared'. Lists of 'characteristic completions' are provided, but again without comment, and there is no statement of rationale, no mention of any definable technique of interpretation other than intuitive empathy, nor any data relating to standardization or to validity, which is assumed. Only passing reference is made to earlier work (e.g. Tendler, Rhode) in the same field. The book is attractively produced, but 40*s.* seems rather much to pay for so fragmentary a presentation of what is not, after all, an original idea.

BORIS SEMEONOFF

Personality in Handwriting. By Eric Singer. London: Gerald Duckworth and Co. Ltd. 1954. Pp. 120. 10*s.* 6*d.*

Graphologie et Physiologie de l'Ecriture. By H. Callewaert. Louvain: E. Nauwelaarts. 1954. Pp. 168+xxxi. 90 Belgian francs.

These two books on graphology offer a remarkable contrast. Mr Singer addresses himself to the wide audience of 'everybody who, up to now, has encouraged my efforts', and expresses the wish that his writings 'will also help to convince wider circles and will bring nearer the day of general official and universal recognition of graphology in this country'. Dr Callewaert, writing for a more specialist public, holds that graphology, understood as a means of judging personality from handwriting, is unscientific and based on no sound statistical or physiological data; the handwriting is not like the expressive gestures and mimicries that accompany speech; and that, like speech transmitted by microphone and so robbed of facial and gestural accompaniments, writing consists largely of laboriously learned and therefore conventional elements.

To psychologists who value the good name of projective methods in the assessment of personality, and who are wary of the 'blind-diagnosis' and the 'dream-book' kind of interpretation, the cold wind of Dr Callewaert's criticism is healthier and more acceptable than Mr Singer's partisan belief that, 'The graphologist can *seldom afford the luxury of interviewing his subject.* It is one of the most significant differences of graphology compared with other personality tests and projection methods that the graphologist does not need the presence of the tested person, and can work without questioning and

interviewing' (p. 55, reviewer's italics). The approach to the subject is dogmatic and unscientific, and the book should appeal only to the uninquiring layman.

Dr Callewaert includes a useful historical summary on writing and graphology and devotes a large part of the book to a consideration of the physiological mechanisms involved in writing. Both books are freely illustrated.

LORNA M. SIMPSON

Die Tests in der klinischen Psychologie (*Tests in Clinical Psychology*). By E. Stern (ed.). Zurich: Rascher Verlag. 1954. Pp. 418. S.Fr. 28.10.

This is the first of two volumes dealing with 'tests in clinical psychology', to which authors of six different countries have contributed. The collection contains twenty-one papers, ten coming from France, five from Switzerland, three from the U.S.A., two from Germany, and one from Portugal. In 20 pages of introduction respectful homage is paid to about thirty scientists predominantly of the medical profession and of little or no historical interest to the problem. In no instance, however, is there reference to any living non-medical experimental psychologist working in the clinical or personality fields.

The book is divided into three parts. The first part contains the introduction, a description of the unstructured interview situation, as well as a historical survey of test development. The second chapter describes intelligence testing, developmental and motor tests, and some elementary statistical methods. Part three introduces the following tests: M.M.P.I. and some other questionnaires, Jung's association test, The Szondi test, Mira's drawing test, and two Swiss tests in which the subject's choice of book titles and proverbs is interpreted.

Most of the contributions represent introductory descriptions of the techniques, and contain little detailed discussion of relevant literature. There is, with the noteworthy exception of the American authors Hathaway and Meehl who deal with the M.M.P.I., little or no factual discussion of test construction, validity, or other statistical requirements. Further drawbacks are the failure to present more objective personality tests of the experimental type, the fact that the bearing of clinical methods on personality theory is not discussed, and that no indication is given of future trends of development.

It is hoped by the reviewer that the few trends towards more rigorous standards found in this publication will not pass unnoticed by the reader to whom it is addressed. The foreign reader can obtain from it a representative account of current continental ideas on clinical testing.

J. BRENGELMANN

Factorial Analysis for Non-mathematicians. By C. J. Adcock. Melbourne: Melbourne University Press. London: Cambridge University Press. 1954. Pp. 88. 17*s*. 6*d*.

The aim of this book is 'to cover the essential processes of factor analysis in such a way as to make clear the underlying logic in simple language'. The approach is through well-chosen numerical examples. Attention is focused on three methods: analysis for simple structure (Thurstone), cluster-directed analysis, and group factors derived from preliminary simple summation (Burt). Criticisms from the point of view of factor analysis are that the basis of factorial methods is not said to be measurements but correlations, that no account is given of how product-moment correlation comes logically to be involved, that use of the tetrachoric coefficient is even made to appear better, and that the mistaken procedure of inserting new communalities for the calculation of each successive centroid factor is advocated. The statistician can also point to the isolation of the discussion from any general consideration of statistical methods and their requirements. The main commentary is contained in chapters on the logic of factor analysis, rotation, interpretation of factors and the role of factor analysis. Here the arguments advanced can be held to suffer from their brevity and may be found persuasive rather than logically compelling. With qualifications, therefore, the book has much to recommend it as a teaching device, and it is of interest for its straightforward presentation and comparisons of the principal methods of group factor analysis.

ARTHUR SUMMERFIELD

Laboratory Instrumentation in Psychology. By W. W. Gringes. Palo Alto, California: The National Press. 1954. Pp. iii + 282. $4.75.

This volume is described as an introduction to the laboratory for graduate students and for professional people who require sufficient detail to enable them to start work in a selected research field. Although the approach is necessarily selective, apparatus of interest to a considerable range of psychologists is

considered. A number of the items described belong to the proverbial 'brass instrument psychology', but their modern electronic counterparts are also included. Each chapter is devoted to a separate field of instrumentation. Special attention should be drawn to the chapters on Audition, Vision and Bioelectricity. The first includes a useful discussion of the physics of sound and also deals in an elementary manner with sound-recording. The second includes a valuable survey of the nature and measurement of light and colour. The third starts with a survey of the types of electrical activity of special interest to the psychologist. The technical problems of electrodes and conducting media suitable for use in examining various forms of electrical activity in the human body are considered. This final chapter ends with an outline of the terminology and basic problems involved in amplifying such electrical phenomena. Each chapter is copiously illustrated with diagrams, photographs and actual electrical circuits where these are relevant. *Laboratory Instrumentation in Psychology* fills an important gap in existing psychological literature.

ROLAND HARPER

Minds and Machines. By W. Sluckin. Harmondsworth: Penguin Books Ltd. 1954. Pp. 223. 2*s.* net.

Mr Sluckin provides, what has not previously existed, a sound, straightforward and accurate account of the body of recent publication and discussion that passes broadly under the heading 'Cybernetics', particularly in its bearing upon psychology. He sustains a high and even level of competence and caution throughout—there is very little that is misleading in this book. Neither is there any false titillation of appetites to which some popularizations of Cybernetics—in common with 'Science fiction'—seem designed to pander. Equally there is little intrepid groping at fundamentals. The result is in many ways a satisfactory, and certainly a useful book.

The principles underlying the representation and communication of information, of mechanical computation, of negative feed-back and of servo-mechanisms are described in terms which are clear and easy to follow. Possible analogies between existing machines and the central nervous system are developed along familiar lines, and these are brought into relationship with some contemporary formulations of psychological problems. It seems a pity that the author hardly hints at the issues raised by 'noise', nor does he provoke the concept of probability to retaliation. Discretion, not valour is his watchword. The book is an excellent two shillings' worth which students of psychology should find will cheer but not inebriate.

R. C. OLDFIELD

L'Esthétique. By Denis Huisman. Paris: Presses Universitaires de France. 1954. Pp. 128. Price not given.

This is a very interesting and illuminating little book, and should have a wide sale and be translated into English. Its author is secretary of the *Revue d'Esthétique* and of the *Société française d'Esthétique*. The book is divided into three parts. In the first part the author deals with the great steps or phases in the history of aesthetics: first, the Platonic or dogmatic age, as he calls it; secondly, the critical or Kantian stage, including pre-Kantian and post-Kantian aesthetics; thirdly the positivistic or modern age. In the second part of the book the author writes about the great fields of aesthetic inquiry: first the philosophy of art; secondly, the psychology; and thirdly, sociology of art. In the third part of the book he deals with special problems and aspects of art and aesthetics: first, the problems of the relationships of art with science, morality, nature, industry and religion; secondly, the classification and grouping of the various arts; thirdly, the aims and methods of aesthetics and its outlook for the future. In all it is a very comprehensive and widely based work, and the author concludes by saying that if aesthetics refuses to be experimental, precise and positive, it will cease to exist.

R. W. PICKFORD

Catholics in Psychology. By Henryk Misiak and Virginia Staudt. MacGraw-Hill Publications Inc. 1954. Pp. xv+309. 40*s.*

This is a book of 300 pages and is one in the series of McGraw-Hill Studies in Psychology. The frontispiece shows a fine portrait of the late Cardinal Mercier, a pioneer in bringing Catholic studies into line with modern psychology, at the University of Louvain. Here was founded a brilliant school of experimental psychology where Michotte became Professor.

The authors start with a good résumé of the rise of psychology in general, pointing out that in Germany there were two main streams, one deriving from Wundt, and the other from Brentano—the founder of 'phenomenology'. The fact that the latter, working in the materialist climate of the nineteenth century, could not reconcile his faith with the new 'science' and gave up being a Catholic, may have been responsible in part for the distrust felt by some Catholic philosophers and theologians for this 'psychology without a soul' which was being torn from the sheltering arms of philosophy.

However, this rift did not long continue, and the cause of free psychology was ably served by Catholic psychologists in various parts of the world, notably Verner Moore—a Benedictine monk and also a psychiatrist, in America—and Agostino Gemelli, Franciscan friar, and also Rector of the University of Milan. Among many brief biographies, there is one on the late Sister Marie Hilda, that remarkable little nun of Glasgow, who started one of the first Child Guidance Clinics in this country, and directed it till her recent death.

'There is of course no Catholic psychology; there is only psychology', as the authors remark, and there is no necessary conflict whatever between the Catholic religion and science.

One wonders whether it was necessary to undertake this lengthy compilation—which is certainly very complete and accurate—in order to demonstrate that there are Catholics working in the psychological field, and that there is now much more understanding on both sides than there was a generation or two ago.

However, anything which breaks down misunderstanding and prejudice is to be welcomed, and the authors have ably carried out their task.

C. L. C. BURNS

Police Work with Juveniles. By JOHN P. KENNY and DAN G. PURSUIT. Springfield, Illinois: Charles C. Thompson; Oxford: Blackwell Scientific Publications Ltd. 1954. Pp. xxi+371. 55*s.* 6*d.*

This book was written as a practical manual for police men and women in the U.S.A. As might be expected part of it is concerned with elementary advice of a factual nature such as how to keep a file and the duties of a policewoman. However, there is much more to the book than this. It does tell the police in a clear and simple way some of the underlying causes of delinquency. Sound advice is also given on how and how not to deal with people, and how and how not to interview. Although the book contains nothing new for psychologists it is an interesting account of a very thorough and praiseworthy attempt to bring the behaviour of the police more in line with the ideas and precepts laid down by psychologists and other experts on delinquency.

MADELINE KERR

Three Men. By JEAN EVANS, with an Introduction by GORDON W. ALLPORT. London: Gollancz. 1954 (first copyright in 1950). Pp. xviii+297. 15*s.*

The author has given three vivid word pictures of delinquent, neurotic or problem boys and of their development. They are Johnny Rocco, probably a psychopathic delinquent, William Miller, probably an hysterical delinquent, and Martin Beardson, a partial homosexual. She describes their personal histories with skilful artistry based on a considerable understanding of clinical psychology, and attempts in a somewhat 'racy' literary manner to make them come to life in her descriptions. In this attempt she is very successful. She undertakes a kind of clinical biography in case-history writing, which would be fascinating and useful to the student of social sciences, of child guidance or the problems of adolescence. Interesting and illuminating though these biographies may be, they are very long and much of the detail in unnecessary. It would have been more useful to have had ten biographies suitably presented in an equally vivid and sympathetic though briefer form. Though the ordinary reader and student will find the novel-like writing very absorbing, he will probably be quite incapable of providing for himself the theoretical interpretation and synthesis which is very necessary and on which the biographies obviously depend though it is never presented. Nevertheless, the book would form an admirable basis for instruction in clinical psychology if the teacher could provide the necessary explanations.

R. W. PICKFORD

The Durham Research Review, No. 5, September 1954. London: University of London Press. Pp. 62. 5*s*.

The contents of this issue are divided between education and psychology. Three articles deal interestingly with byways in English educational history. J. Conway Davies describes, with frequent quotation from his sources, some Elizabethan schemes of education, and J. L. Dobson gives a parallel detailed picture of a school run on Dr Bell's system in the early nineteenth century. The feature of both schemes is the exacting demands on the pupil, and the apparent absence of any attempt to fit the curriculum to the pupil's interests and abilities. An article by M. Gilbert discusses the influence of the German reformers of the 1870's on language teaching in England, illustrating incidentally the transition from the educational attitudes of Dr Bell's time to that of the present day. All these articles are adequately documented.

Of more interest to the readers of this *Journal* will be the psychological contributions. Dr Macfarlane Smith describes the construction, analyses and validation of a Space Test. He confirms the evidence for a 'Space Factor' at 11+, and finds the space test correlates more highly with school marks in Technical Subjects than does the conventional verbal group test. But the correlations are in general low, and the numbers not large. Dr Smith also speculates on the relationship of space-test scores to masculine and feminine traits. T. P. Howard, in the field of experimental psychology, examines the conditions under which the size-weight illusion operates, finding a change from positive to negative error as the weights used become smaller. The discussion raises a number of theoretical points which could profitably be pursued farther. Graham and Sluckin report on a pilot study of factors which workers consider important in their jobs. The inquiry is slight, but does not claim to be more.

The difficulty of editors of such periodicals as this is to maintain a supply of articles of adequate quality in competition with the more specialized journals. The heavy weights are normally attracted to the latter type of journal; but Prof. Smith and his co-editors can be congratulated on a set of competent and interesting light-weight contributions.

J. MAXWELL

The Teaching of Arabic Numerals. By G. G. Neill Wright. The Scottish Council for Research in Education. XXXVI. The University of London Press. 1954. Pp. 45. 2*s*. 6*d*.

This is a summary of *The Writing of Arabic Numerals* by the same author published in 1952 and reviewed in this *Journal* in February 1954. It is published by the Scottish Council for Research in Education in pursuance of their policy of making the results of investigations easily available to those who are more interested in the results themselves than in the steps by which they have been reached. Readers who wish to review the evidence on which the results rest are referred to the larger volume.

This very much abridged version is a very successful summary. It brings out clearly the importance of teaching the child a form or prototype for each numeral which he can write legibly, quickly, with ease, and as nearly as possible independent of attention. Advice on the selection of prototypes is given and many illustrations are provided.

For any one concerned with teaching children arithmetic this is a most useful little book.

JOHN SUTHERLAND

OTHER PUBLICATIONS RECEIVED

A. M. Blackwell. *Lists of Researches in Education and Educational Psychology: Supplement I.* (Publications of the National Foundation for Educational Research in England and Wales, No. 7.) London: Newnes Educational Publishing Co. Ltd. 1954. Pp. 57. 10*s.* 6*d.*

Albert Burloud. *Psychologie de la Sensibilité.* Paris: Libraire Armand Colin. 1954. Pp. 224. 250 francs.

Egon Freiherr von Eickstedt. *Atom und Psyche.* Stuttgart: Ferdinand Enke. 1954. Pp. 158. Paper, DM. 12.00; linen, DM. 14.20.

Peter Fireman. *Perceptualistic Theory of Knowledge.* New York: Philosophical Library. 1954. Pp. xi+50. $2.75.

A. Michotte. *La Perception de la Causalité.* Anvers (Belgium): S. A. Standaard-Boekhandel; Second edition, 1954. Pp. viii+306. Paper, 175 Belgian francs; bound, 220 Belgian francs. (A reprint of the original text, with a new 2-page preface and a short list of relevant research published since the appearance of the first edition (1946).)

John Penman and Marion C. Smith. *Intra-Cranial Gliomata.* (Med. Res. Council, Special Reports series, no. 284.) London: H.M.S.O. Pp. vi+70. 1954. 5*s.*

Calvin P. Stone. *Abnormal Psychology Glossary* (revised edition). Stanford University Press; London: Geoffrey Cumberlege. 1954. Pp. 24. 4*s.* net.

British Film Institute. *Films on the Sciences.* London: British Film Institute, July 1954. 1*s.*

J. Campbell (ed.). *Spirit and Nature: Papers from the Eranos Yearbooks.* London: Routledge and Kegan Paul. 1955. Pp. xvi+492. 36*s.*

Mirclea Eliade. *The Myth of the Eternal Return.* (Translated from the French by W. R. Trask). London: Routledge and Kegan Paul. 1955. Pp. xi+195. 18*s.*

J. J. Poortman. *Ochêma (History and Import of Hylic Pluralism).* Assen, Holland: Van Gorcum. Pp. 179 (incl. 8-page Synopsis in English). Paper Fl. 6.90, cloth Fl. 8.25.

W. Bromberg. *Man above Humanity: a History of Psychotherapy.* London: Pitman Periodical Publishing Co. 1954. Pp. xiii+342. 45*s.* net.

D. McLean. *Nature's Second Sun: Leaves from a Teacher's Log.* London: Heinemann. 1954. Pp. xii+130. 10*s.* 6*d.* net.

J. C. Flugel. *Man, Morals and Society.* Harmondsworth: Penguin Books Ltd. 1955. Pp. 415. 3*s.* 6*d.* (With a new Preface.)

J. B. Rhine. *The Reach of the Mind.* Harmondsworth: Penguin Books Ltd. 1954. Pp. 198. 2*s.* 6*d.*

G. Cuisenaire and C. Gattegno. *Numbers in Colour: A new Method of Teaching Arithmetic in Primary Schools.* London: Heinemann. 1954. Pp. viii+39. 4*s.* 6*d.*

A. A. Congalton (ed.). *Hawera—a Social Survey: A report of a Community Venture.* Hawera, New Zealand. 1954. Pp. xv+219. 10*s.*

A. J. Nixon. *Divorce in New Zealand: a Preliminary Survey.* Auckland University College (Bulletin No. 46, Sociology Series no. 1). 1954. 4*s.*

Charles Eisenmann. *The University Teaching of Social Sciences: Law.* UNESCO. 1954. Pp. 133. 6*s.* net.

EDITORIAL NOTE

Intending contributors are asked to note that changes have been made in the instructions on the *Preparation of Papers for the British Journal of Psychology* which appear on the inside front cover.

The principal alteration is that it is recommended that in future each article should be preceded by a short abstract or summary, conveying the gist of the article and embodying the principal conclusions. This should, to some extent, render the inclusion of a formal summary on a similar scale at the end of an article redundant, although a concluding section or paragraph summarizing the argument or drawing attention to the main findings will not be out of place. The guiding principle should be to regard the summary as being quite separate from the article proper, which in its turn should not depend in any way for comprehensibility or completeness on the summary.

The summary, placed at the beginning of the article, will also replace the list of numbered headings and sub-headings which at present appears in this position. It is felt that such a list is unnecessary for articles of a length falling within the present recommended maximum. Articles should, however, still be subdivided in the way described.

It is hoped in future to publish a small proportion of very short articles (under 1000 words), such as Apparatus Notes, Research Reports, etc. It may occasionally be possible, as space allows, to accord priority to such contributions. Short papers of this kind may not require to be subdivided; a summary may also be unnecessary.

All book reviews will in future be signed. Reviewers are reminded that full bibliographical data should be given, in the following order: Title; author; place of publication; publisher's name; date; number of pages (with preliminary pages noted separately); price. Where a publisher acts as agent for a foreign house this should be mentioned.

The use of roman numerals in the citation of volume number references is being discontinued.

The new style will come into effect with Volume 47. Since it will not be possible to include in the present volume articles other than those already received, all contributions submitted from now on should conform to the altered requirements.

ERRATUM

The authors of the article "The effect of hunger upon drinking patterns in the rat" (A. D. Calvin and R. A. Behan, this *Journal*, XLV, Part 4) have intimated that the following footnote to their title was inadvertently omitted:

"The authors are indebted to E. F. MacCaslin whose observations prompted this experiment."